# HYDRODYNAMIC ANALYSES OF SURFACE WATER FLOW

A thesis presented for the degree of

Doctor of Philosophy

in

Civil Engineering

in the University of Canterbury,

Christchurch, New Zealand

by

A. G. Barnett

1970

ACKNOWLEDGEMENTS

I wish to acknowledge the assistance given me by the following:

Professor H.J.Hopkins, who as head of the Civil Engineering Department helped to create the environment which made this thesis possible.

Dr A.J.Sutherland for guidance during the projects, and particularly for many constructive suggestions concerning this thesis.

Professor F.M.Henderson for supervising the early stages of this project and for first inspiring my interest in hydraulics.

Other members of the academic staff of the University of Canterbury.

The technical staff of the Civil Engineering Department.

The Ministry of Works, for financial support during the project.

Dr J.V.Robinson and Dr S.M.Thompson of the Systems Laboratory, Ministry of Works, for their support which facilitated the completion of this thesis.

Mrs R. Brennan for her painstaking and expert typing.

Finally I wish to dedicate this thesis to my wife Jane, who has made appreciable sacrifices to assist and encourage me in completing my project.

## ABSTRACT

A general system of one dimensional overland flow equations, with provision for laminar or turbulent flows, mildly irregular surfaces, lateral inflows and steep slopes is derived. The failure of existing solution theory to predict the reliability of a difference solution of such equations is traced to inadequacies in the treatment of non-homogeneous terms. A new approach to Fourier series stability analysis based on linearising variations in the non-homogeneous terms rather than neglecting them is developed, and this resolves the apparent discrepancy between the predictions of numerical analysis and the actual behaviour of numerical solutions. Stability properties deduced by this analysis include known physical flow stability criteria. A difference scheme proposed by Courant, Isaacson and Rees is modified to a semi-explicit form, rectifying its stability defects. This Semi-Explicit scheme is shown to possess advantages over characteristic, implicit and explicit schemes, and a method is given for choosing efficient, stable time steps for numerical solutions.

The value of simplified solutions in special cases is illustrated by the successful prediction, by diffusion methods, of the properties of an artificially produced flood in a large river.

## PREFACE

Any synthesis which is aimed at predicting certain features of a physical system, such as an overland flow, from other properties of that system, can be divided into four interdependent stages - analysis, measurement, solution and application. This thesis covers research conceived as part of a project investigating the interrelationship of these four stages with the final aim of establishing the validity of numerical simulation of overland flow by direct comparison between computed numerical solutions and experimental results from a simple laboratory catchment. Thus although this thesis is principally concerned with the analysis and solution stages which deal with an idealised representation of the actual system, an important consideration has been the effects of inconsistencies between measured physical properties and their idealised form, for instance, the inconsistency between the actual bed geometry of a natural channel and a single number representing "bed slope". These effects can be minimised by simulating the overland flow system as comprehensively as possible, and every effort has therefore been made to restrict mathematically convenient simplifying assumptions to those essential to the argument.

The project investigating all stages of the synthesis proved to be too ambitious because current solution theory proved to be less adequate than a superficial survey of the literature to 1966 had suggested; in particular none of the contemporary methods of investigating stability seemed to take account of the obvious instabilities discussed in Section 4.2. Extended Fourier series stability analysis was then developed, essentially in the form presented in Chapters 5 and 6. This work was reported to the University of Canterbury in March 1968, which is the reason why the recent valuable stability analyses given by Vreugdenhil (1968) and Liggett and Woolhiser (1968) are not incorporated in the presentation to the extent which they perhaps deserve.

Despite the delays to the original project resulting from the time spent on stability analysis methods, the partly completed experimental programme was continued in an attempt to fulfil the original aim of comparing numerical solutions and experimental results, but such comparisons proved to be too complex to permit adequate treatment in the time remaining. It was therefore decided to concentrate on the development of the Semi-Explicit difference scheme and to postpone further analysis of the experimental results until after the completion of this thesis. Thus it is hoped to make more adequate use of the contribution to the project by the technical staff of the Fluid Mechanics laboratory, P.C.Dawson, G.F.Archer and in particular H.S.Pearce, at a later date.

A.G.Barnett

CONTENTS

## PART II FINITE DIFFERENCE SOLUTIONS

### Chapter 3

### EXISTING SOLUTION THEORY

Chapter 5

## LIMITED STABILITY TESTS ON A GENERAL FINITE DIFFERENCE SCHEME



Chapter 6

## COMPLETE STABILITY ANALYSIS ON SELECTED FINITE DIFFERENCE SCHEMES

## LIST OF FIGURES



## LIST OF TABLES

## LIST OF SYMBOLS

It has not been possible to avoid repeated use of many symbols for different purposes because of the volume of algebraic manipulation in this thesis, but each symbol is defined where it first appears with a new meaning, and different uses of the same symbol are separated as widely as possible. As noted in Section 2.2, the subscripts t, s, x or y define partial derivatives, while any other subscript simply distinguishes between similar variables. An asterisk * as a superscript refers to a footnote except with $F_f^*$ and $F_R^*$ in Sections 2.3, 2.7. A superior asterisk (e.g. $\overset{*}{u}$) denotes a Fourier transform as defined in Appendix A, while a superior stop (e.g. $\overset{\circ}{V}$) denotes a small variation quantity. Vertical enclosing lines denote the determinant of a matrix in Chapter 3, but elsewhere (see Section 5.7) they denote the modulus of a scalar. The transformation symbol $\Longrightarrow$ is defined in Section 5.6. Any superscript 0, 1, ..... b, b+1, ..... refers to the value of the variable at time (0, 1, .... b, b+1, ....)$\Delta t$ and if an associated subscript a-1, a, a+1 appears the variable is also evaluated at the point (a-1, a, a+1)$\Delta s$. All subscripts not involving t, s, x, y or a$\Delta$s, and all superscripts not involving b$\Delta$t, are defined with their associated symbols in the following list.

| | |
|---|---|
| $A$ | Cross-sectional area of flow, Figure 2-1 |
| $A_{P1,2,3}$ | Projected cross-sectional areas, Figure 2-2(b) |
| $a$ | "Distance" subscript explained above, introduced in Appendix A |
| $B$ | Flow width at surface, introduced in Section 2.8 |
| $b$ | Barometric pressure, Section 2.3<br>"Time level" subscript introduced in Appendix A |
| $C$ | Constant in turbulent friction slope, defined in (2-39) or (2-41) |
| $C_{1,2}$ | Constants of integration, Section 2.9 |
| $C_1$ | Constant, Appendix A |
| C.R. | Criterion Ratio defined by (7-30) |
| $c$ | Wave celerity in Part II, defined in (3-14)<br>Wave "speed" in Part III, defined in (8-4) |
| $c^1$ | Fitted wave "speed", Section 9.4 |
| $c_m^b$ | Mean value of $c_a^b$, defined in (4-8) |
| $D$ | Generalised constant in friction slope formula (2-53) |
| $D_c$ | Difference coefficient, defined in (A-9)<br>Difference coefficient, defined in (7-11) |
| $D_V$ | Difference coefficient, defined in (7-10) |
| $D_v$ | Difference coefficient, defined in (A-8) |
| $d$ | Differential operator<br>Mesh refinement ratio, Section 3.5 |
| $E$ | Simple channel coefficient, defined in (2-37) |
| $E_{0,1,2...}$ | Coefficients of $y^{0,1,2...}$ in area/depth polynomial (3-3) |
| $e$ | Depth of channel axis, Figure 2-1<br>The exponential number, 2.71828.. except in Chapter 2 |
| $F$ | Froude Number in Part II, defined in (3-18)<br>A function defined in (8-10)<br>Input function, Appendix F |

| | |
|---|---|
| $F_{1,2}$ | Functions in Section 8.2 |
| $F_c$ | "c factor" in Section 8.5 |
| $F_f$ | Local shear force, introduced in (2-7) |
| $F_f{}^*$ | Sum of $F_f$ |
| $F_K$ | "K factor" in Section 8.5 |
| $F_R$ | Local normal force, introduced in (2-7) |
| $F_R{}^*$ | Sum of $F_R$ |
| $F_w$ | Local weight force, introduced in (2-7) |
| f | Section ordinate, Figure 2-1<br>Variable, introduced in (8-8) |
| $\mathring{f}(s)$ | Any locally constant variation quantity, introduced in Section 5.5 |
| $f_{1,2...}$ | Solution functions, Appendix F |
| G | Variable, defined in (3-16)<br>Dimensionless channel parameter, defined in (F-13) |
| $G_{1,2,3,4}$ | Coefficients representing flow properties, defined in (6-7) to (6-10) |
| $G_{t5,t25..}$ | G associated with $t_5$, $t_{25}$.... in Table 9-1 |
| g | Acceleration of gravity, Section 2.1 |
| H | Horizontal axis, Figure 2-1<br>Shape parameter in Part II, defined in (3-2)<br>Dimensionless channel parameter, defined in (F-14) |
| $H_{50}$ | Value of H associated with $t_{50}$, Section 8.5 |
| $H_p$ | Value of H associated with $t_p$, equation (8-23) |
| $H_{p1}, H_{p2}$ | Value of H associated with $t_{p1}, t_{p2}$ |
| h | Piezometric head above a horizontal datum, Figure 2-3<br>Coefficient in Appendix C, defined in (C-3) |
| $h_0$ | Initial hydraulic head, Section 9.4 |
| I | Signifies independent relationship with some of s,t,V,A,R,y in Chapter 2<br>Identity matrix in Appendix A and Part II |

| | |
|---|---|
| $I_{1-5}$ | Real coefficients of imaginary terms, Appendix B |
| $i$ | Velocity profile parameter, defined in (2-21)<br>$\sqrt{-1}$, Appendix A, Part II |
| $J$ | Coefficient, Section 2.3<br>Diffusion coefficient, equation (F-2) |
| $J_{1,2}$ | Used for condensing algebra in Appendix B. Defined in (B-13) and (B-16) |
| $j$ | Slope exponent in (2-54) |
| $K$ | Diffusion coefficient defined in (8-5) or (F-2) |
| $K^1$ | Fitted diffusion coefficient, Section 9.4 |
| $K_{1,2}$ | Constants, equation (4-10) |
| $K_{a,b}$ | Limits on a,b in Appendix A |
| $K_{a,b,c}$ | Coefficients of quadratic, defined in (7-23) to (7-25) |
| $k$ | Dimensionless cross-sectional geometry parameter, defined in (2-35) |
| $L$ | Dimensionless cross-sectional geometry parameter, defined in (2-36)<br>Limit O(1/10), Appendix C |
| $L_1$ | Coefficient representing flow properties, defined in (6-59)<br>Special case of above (except for a factor of 2), defined in (4-37) |
| $L_2$ | Coefficient representing flow properties, defined in (6-60)<br>Special case of above (except for a factor of 2), defined in (4-37) |
| $M$ | Constant in (2-41) |
| $M_{1,2,3}$ | Matrices in Section 5.7 |
| $M_a$ | Amplification matrix, introduced in (A-13) |
| $m$ | Hydraulic radius exponent in (2-54)<br>Typical term in the series in (C-3) |
| $N$ | Normal axis, Figure 2-1, also dimension along normal axis, Section 2.9<br>Convenient substitution in Appendix B, defined in (B-6) |

| | |
|---|---|
| $n$ | Manning's n, defined in (2-41)<br>Number of operations, Appendix A<br>Typical term in the series in (C-2) |
| $n_{1,2,3,4}$ | Real constants in Section 5.9 |
| $P$ | Wetted perimeter, defined in Section 2.8<br>Point label, Section 3.4, Appendix A |
| $P'$ | Point label, Section 3.4 |
| $P_{1,2,3}$ | Coefficients defined in (6-94), (6-95), (6-96) |
| $P_{4,5}$ | Coefficients used in Section 6.12 to relate $P_2$ and $P_3$ |
| $p$ | Local lateral inflow, Figure 2-1<br>Fourier series parameter introduced in (A-12)<br>Convenient substitution defined in Section 7.4<br>Percentage change in stage, defined in (8-22) |
| $p_{1,2}$ | Given percentage changes in stage, introduced in Section 8.4 |
| $Q$ | Channel discharge, Figure 2-1 |
| $q$ | Lateral flow to channel per unit length, introduced in Section 2.2 |
| $R$ | Hydraulic radius defined in (2-31) |
| $R_{1-5}$ | Coefficients of real terms, Appendix B |
| $r$ | Horizontal length of isotach, Figure 2-5(a)<br>Ratio $\Delta t/\Delta s$ in Part II, Appendix A |
| $S_f$ | Friction slope, defined in (2-13) and evaluated in (2-53) |
| $s$ | Distance along the channel axis, introduced in Section 2.1 |
| $s_1$ | Downstream boundary point in Appendix F |
| $T$ | Wind shear in Section 2.9<br>Time of solution, introduced in Appendix A |
| $t$ | Time ordinate, introduced in Section 2.1 |
| $t_{5,50..}$ | Time elapsed until 5%,50%... change. Introduced in Section 8.5 |

$t_o$ Initial time in Section 3.4

$t_{p,p1...}$ Time elapsed until $p\%, p_1\%...$ change. Introduced in Section 8.4

U Average relative velocity of lateral flow, defined in (2-10)

u $s$ component of local lateral flow velocity, Figure 2-1
Vector in Appendix A

V Mean velocity defined in Section 2.3

$V_{1,2}$ Measured velocities, Section 3.2

$V_m^b$ Mean value of $V_a^b$, used in Section 4.2

$V_o$ Uniform flow velocity, used in Section 5.4

$\overset{+}{V}_s, \overset{-}{V}_s$ Partial derivatives associated with the forward and backward characteristics respectively, introduced in Section 5.6

v Local velocity, Figure 2-1

W $V\Delta t/\Delta s$, defined in (A-18)

w Vertical ordinate of the s axis, Figure 2-3
"Stage variable" defined by (3-22)

$\overset{+}{w}_s, \overset{-}{w}_s$ Partial derivatives associated with the forward and backward characteristics respectively, introduced in Section 5.6

X Ratio defined in (2-47)
Convenient substitution in Chapter 6 and Appendix B, defined in (6-49)

x Horizontal component of s, Figure 2-1
Parameter in Appendix C, defined in (C-1)

$x_1$ Solution of x in (C-10)

Y Any steady stage, Section 8.2

$Y_0$ Initial steady stage in Part III, Appendix F. Assumed constant in Appendix F only.

y Distance in -N direction from datum to water surface. Introduced in Section 2.7

| | |
|---|---|
| $y_0$ | Final change in stage in Part III, Appendix F. Assumed constant in Appendix F only |
| $y_{1,2}$ | Measured stages, Section 3.2 |
| Z | $c\Delta t/\Delta s$, defined in (A-18)<br>"Dummy variable" defined in (F-11) |
| z | Vertical ordinate of datum, introduced in Section 2.7 |
| $\alpha$ | Convenient substitution in Appendix F |
| $\alpha_{1,2}$ | Coefficients related to the point of evaluation of the non-homogeneous terms, introduced in Section 6.1 |
| $\beta$ | Convective acceleration corrective coefficient, defined in (2-14) |
| $\Gamma_{1,2}$ | Convenient substitutions defined in (6-61), (6-62) |
| $\gamma$ | Large positive constant, Section 3.5<br>Lateral flow parameter defined in (4-33), special case of $\gamma_1$, $\gamma_2$<br>Convenient substitution, Appendix F |
| $\gamma_{1,2}$ | General lateral flow parameters defined in (6-5), (6-6) |
| $\Delta$ | Small increment of .. |
| $\Delta t_{1,2}$ | Critical time steps for the CIR, Semi-Explicit Schemes, introduced in Section 7.7 |
| $\delta\phi$ | Small error in $\phi$, Section 2.4 |
| $\partial$ | Partial differential operator |
| $\epsilon$ | Used in Sections 4.6 and 4.7, and Appendix B, as one coefficient in the unknown square root, see (B-1)<br>Error in Appendix C, defined in (C-4) |
| $\epsilon_{1,2}$ | Alternative coefficients of the square root |
| $\epsilon_r$ | The real alternative, $\epsilon_1$ or $\epsilon_2$ |
| $\zeta$ | Used in Sections 4.6 and 4.7, and Appendix B, as the other coefficient in the unknown square root, see (B-1) |
| $\zeta_{1,2,r}$ | Corresponds with $\epsilon_{1,2,r}$ respectively |

| | |
|---|---|
| $\eta$ | Local acceleration corrective coefficient, defined in (2-15) |
| $\theta$ | Substituted for $p\Delta s$, introduced in Appendix A |
| $\kappa_{1,2}$ | Coefficients with modulus no greater than unity. Defined in (6-77) |
| $\Lambda_{1,2}$ | Eigenvalues of $M_3$, introduced in (5-42) |
| $\lambda$ | Either eigenvalue of the amplification matrix, introduced in Section 4.6. "Dummy variable" in Hayami's solution (F-5) |
| $\lambda_{1,2}$ | Eigenvalues of the amplification matrix in Appendix A and Chapters 5 and 6 |
| $\lambda_j$ | Eigenvalues of the general amplification matrix, Appendix A |
| $\mu$ | Dynamic viscosity, introduced in (2-43) |
| $\mu_{11,12..}$ | Elements in $M_3$, defined in (5-43) |
| $\pi$ | Circular circumference-diameter ratio, 3.14159... |
| $\rho$ | Mass density, introduced in Section 2.1 |
| $\Sigma$ | Summation symbol, first used in (2-10) |
| $\sigma$ | Friction slope parameter, special definition (4-33), general case (6-4) |
| $\tau$ | Shear stress, Figure 2-5(c) |
| $\tau_{1,2,3,4}$ | Coefficients, defined in (5-30), related to the non-homogeneous terms |
| $\Upsilon_{1-8}$ | Elements of $M_1$ and $M_2$, used for convenience in Section 5.7 |
| $\upsilon_{1,2,3,4}$ | Coefficients, defined in (5-31), related to the non-homogeneous terms |
| $\Phi_{1,2}$ | Convenient substitutions defined in (5-38), (5-39) |
| $\phi$ | Angle between s and x axes, Figure 2-1 |
| $\chi_{1,2}$ | Coefficients, defined in (5-24), related to the time differencing |

$\Psi_{1,2}$ Convenient substitutions defined in (5-36), (5-37)

$\psi_{1,2,3,4}$ Coefficients, defined in (5-26), (5-27), related to the space differencing

$\Omega$ Convenient substitution defined in (5-45)

# HYDRODYNAMIC ANALYSES OF SURFACE WATER FLOW

## PART I - OVERLAND FLOW EQUATIONS

## CHAPTER 1

## INTRODUCTION

### 1.1 Context

Surface water flow frequently assumes considerable importance in engineering schemes. Quantitative descriptions are therefore required of a variety of surface flows, from thin sheet runoff over airport runways to fluctuating turbulent flows in power canals. All these flows are commonly treated by a one dimensional analysis which uses the continuity equation and some relationship between the forces acting and the flow properties to solve for the dependent variables velocity and depth. When this relationship takes the form of Newton's second law of motion the analysis is soundly based on the principles of hydrodynamics and results in an hyperbolic system of partial differential equations which we shall call overland flow equations. An example of a system of overland flow equations is that derived by Stoker (1957) p.455.

Because such an hyperbolic system of equations can only be solved by tedious numerical methods, the normal practice has been to adopt a variety of simplifications to suit individual problems, with corresponding errors in the solutions. Simplifications do offer valuable insights into the behaviour of a solution, and in special cases simplified solutions may be shown to be adequate by experimental verification, but in general simplified solutions are an inferior substitute for the ideal of accessible solutions of general systems of overland flow equations. The advent of reliable computers has now made heavy numerical calculation more practicable, but many users of finite difference methods to solve overland flow equations have reported unexpected difficulties. For instance, Stoker (1957) p.490 was forced to use much smaller finite difference intervals in some parts of his solution than in others, while Fenzl (1965) and Liggett and Woolhiser (1967) reported numerical instability in some calculations. These difficulties all arose in calculations using finite difference schemes which are apparently "theoretically stable" according to current methods of stability analysis. Such methods are usually either similar to that presented by Courant, Isaacson and Rees (1952) or are based on the Fourier series methods introduced by von Neumann, and developed by Richtmyer (1957).

Thus present experience with numerical solutions of overland flow equations suggest an inadequacy in current numerical stability theory which is hindering the exploitation of computers in hydraulic design.

## 1.2 Thesis Outline

This thesis presents a new Fourier series stability analysis which, when applied to a general system of overland flow equations, resolves the apparent discrepancy between the predictions of theoretical stability analysis and the results of numerical experiments. This opens a wide range of overland flow problems to a dependable general purpose numerical solution, but an example is given of a special case where a solution of simplified equations is still of considerable value.

The thesis is accordingly divided into four parts, while additional material which is parenthetical to the main argument is placed in six appendices.

Part I Overland Flow Equations

Likely occurrences of overland flows in engineering design problems are indicated and overland flow equations are introduced. The scope of the project is then defined briefly by reference to the literature, and the organisation of the thesis is outlined. In Chapter 2 equations of continuity and motion are derived by the standard one dimensional treatment of near parallel flow, but as overland flow may occur in a wide range of conditions every effort is made to restrict simplifying assumptions to those necessary to make the analysis tractable. Thus the equations presented at the end of Chapter 2 are a general system of overland flow equations which can be called the Overland Flow equations, and these are used as the basic analytic description of water flows in the remainder of the thesis.

Part II Finite Difference Solutions

The Overland Flow equations are reduced to characteristic form and the influence of the properties of a wave propagation system on the method of solution are indicated in Chapter 3.

Consistency, convergence and stability are defined, following Richtmyer (1957), and the important work of Lax, Courant, Isaacson and Rees on these subjects is summarised.

In Chapter 4, cases where Richtmyer Fourier series stability analyses have proved to be inadequate are cited from the literature and it is shown that in some cases a numerical solution will clearly be unstable even though Richtmyer analysis suggests stability. Richtmyer analysis therefore applies in ideal conditions which are not necessarily met in practice, and a more practical approach to stability is seen to be required. A new approach to Fourier series stability analysis is then established by a relatively simple example. In Chapter 5 the scope of this new analysis is widened to all simple implicit or explicit regular net difference schemes for solutions of flows in regular or mildly irregular channels. A number of general stability criteria for all such schemes are derived. These criteria are interpreted in physical terms in Chapter 6, and one set are shown to correspond with known criteria for physical flow stability, the first time that the axiomatic link between physical and numerical stability has been demonstrated by numerical analysis. Other physical stability criteria are also derived.

Chapter 6 concludes with a complete stability analysis, within the assumptions of Chapter 5, of selected difference schemes and one, the Semi-Explicit scheme, is found to have the best stability properties. In Chapter 7 this scheme is shown to be superior to all of a wide range of difference schemes for general use, and its exact formulation is set out. A series of numerical experiments are then shown to consistently corroborate predictions of the new stability theory and to confirm the advantages of the Semi-Explicit scheme.

Part III The Diffusion Analogy Simplification

The extra assumptions involved in the Diffusion Analogy simplification of the Overland Flow equations are discussed in Chapter 8. It is shown that the diffusion solution given by Hayami (1951) and outlined in Appendix F, can be extended to non-prismatic channels which permits more realistic simulation of floods in natural channels. Attention is concentrated on waves produced by the operation of artificial controls, and it is shown that diffusion theory predicts constant travel times for a given part of any such wave. These predictions are verified in Chapter 9 by the close correspondence between the actual travel times of an artificially generated flood wave on the Clutha River, N.Z. and those predicted from the travel times of another flood with a different discharge/time relationship.

Part IV Summary and Conclusions

Basic assumptions are discussed and conclusions reached in the thesis are summarised.

# CHAPTER 2

# THE ANALYSIS OF SURFACE WATER FLOW

## 2.1 Definition of the Problem

We now proceed to discuss the derivation of the Overland Flow equations, which are the basic general analytic description of water flows to be used in this thesis. These equations, which may be applied to many surface water flows, are a convenient representation of the continuity principle and Newton's second law of motion. They are an extension of the St. Venant equations (Ven te Chow (1959) p.528) which are derived for turbulent flows in near-horizontal channels. We use a one-dimensional analysis, assuming that the water flows in a single stream tube with a reasonably straight axis, the s axis. Our definition of a stream-tube differs from the usual definition in that lateral flows through the walls are permitted, but such flows are identified separately.

It is assumed that the streamlines are approximately parallel to the s axis and that their curvature is small, which corresponds to the assumption of "long" or "shallow water" waves in two dimensional wave theory (Milne-Thomson (1960) p.414, Rouse (1950) p.718). If the lateral inflow has negligible velocity normal to the s axis, this assumption means that there are no appreciable accelerations normal to the s axis and that the velocities normal to the s axis are negligible. This in turn means that in any section normal to the s axis, the piezometric pressure is constant and the free surface will be horizontal.

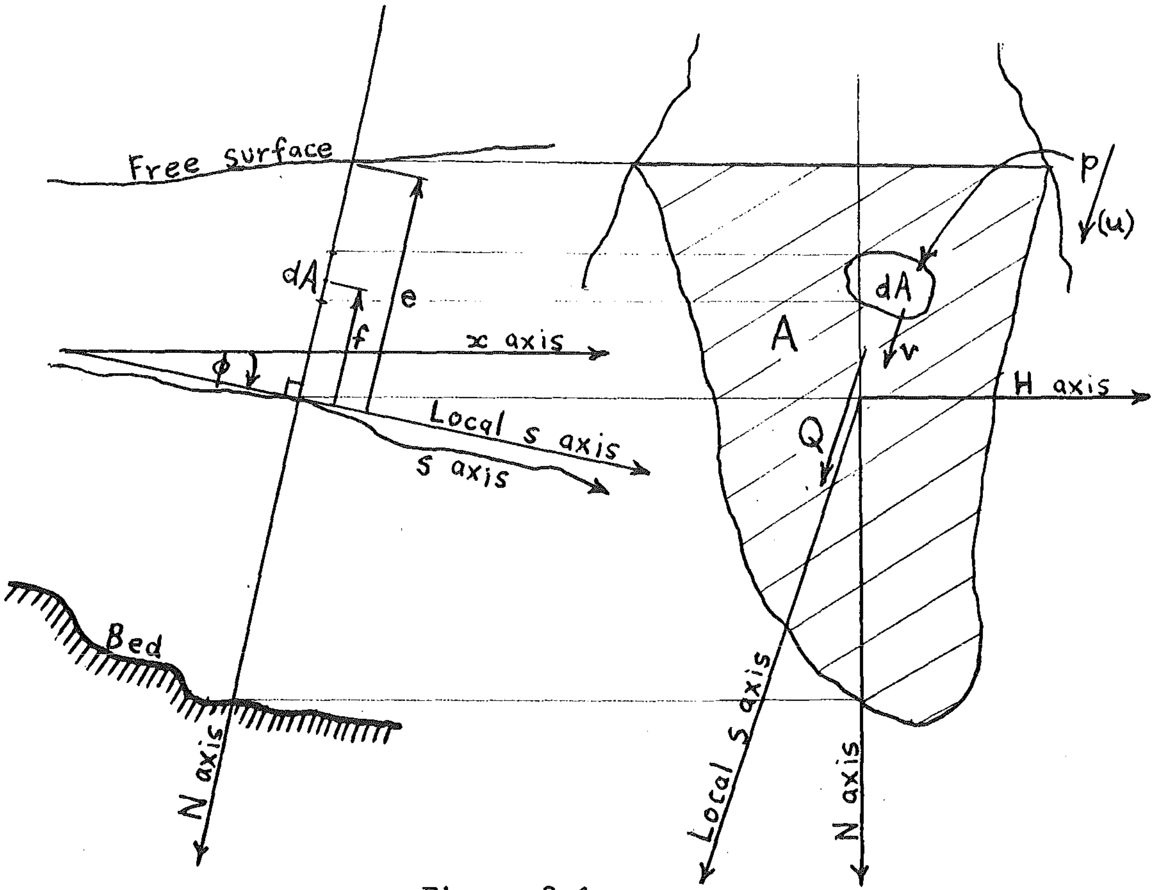



Figure 2-1

Referring to Figure 2-1, we have a streamtube cross-section of area A with a discharge of Q passing through it. The velocity through any small area dA is v, and a lateral inflow of p enters a small elemental streamtube of cross-sectional area dA and unit length with a velocity component u parallel to the s axis. We define a cartesian set of local axes s, N, H as shown in Figure 2-1, where the H axis is horizontal. We add an x axis, which is the projection of the s axis on a datum horizontal plane. The angle between the x and local s axes is $\phi$(+ve downwards). Let the distance of the free surface above the s axis measured parallel to the N axis be e, and the similar distance of dA above the s axis be f.

We use t for time, $\rho$ for the density of water (assumed constant), and g for the acceleration of gravity throughout this thesis.

Since the piezometric pressure is constant in any cross-section

(Gauge Pressure at element dA) = $\rho g(e-f) \cos \phi$ (2-1)

In the standard derivation of the St. Venant equations, three further assumptions are made at this point (Stoker (1957) Pp.452-455).

(a) $\phi$ is assumed to be small so that $\cos \phi \approx 1$.

(b) The use of one dimensional analysis is taken to include the assumption that the actual flow is well approximated by a flow with uniform velocity over any cross-section.*

(c) The lateral inflow is assumed small compared with the main flow Q.

As we wish to analyse overland flows, a possible case is a laminar type flow over a steep hill slope subject to heavy rain. None of the above three assumptions apply to this flow, so that different assumptions must be made when required. The remainder of the analysis therefore, while similar to the derivation of the St. Venant equations, leads to an extension of

---

* Uniformity of velocity over a cross-section is often regarded as a direct consequence of hydrostatic pressure distribution (Lamb (1945) p.254). However this follows only if the forces tangential to the streamtube are neglected (Lamb (1945) p.1).

the St. Venant equations, the Overland Flow equations.

## 2.2 The Continuity Equation

A length of streamtube bounded by two faces normal to the s axis and $\Delta s$ apart, $\Delta s$ being small, is now considered (see Figure 2-2(a)). The continuity law equates rate of increase in mass to the mass inflow into an element, so, following a small element of cross-sectional area dA along its pathlines, we get

$$\frac{d(\rho dA\Delta s)}{dt} = \rho p\Delta s \qquad (2\text{-}2)$$

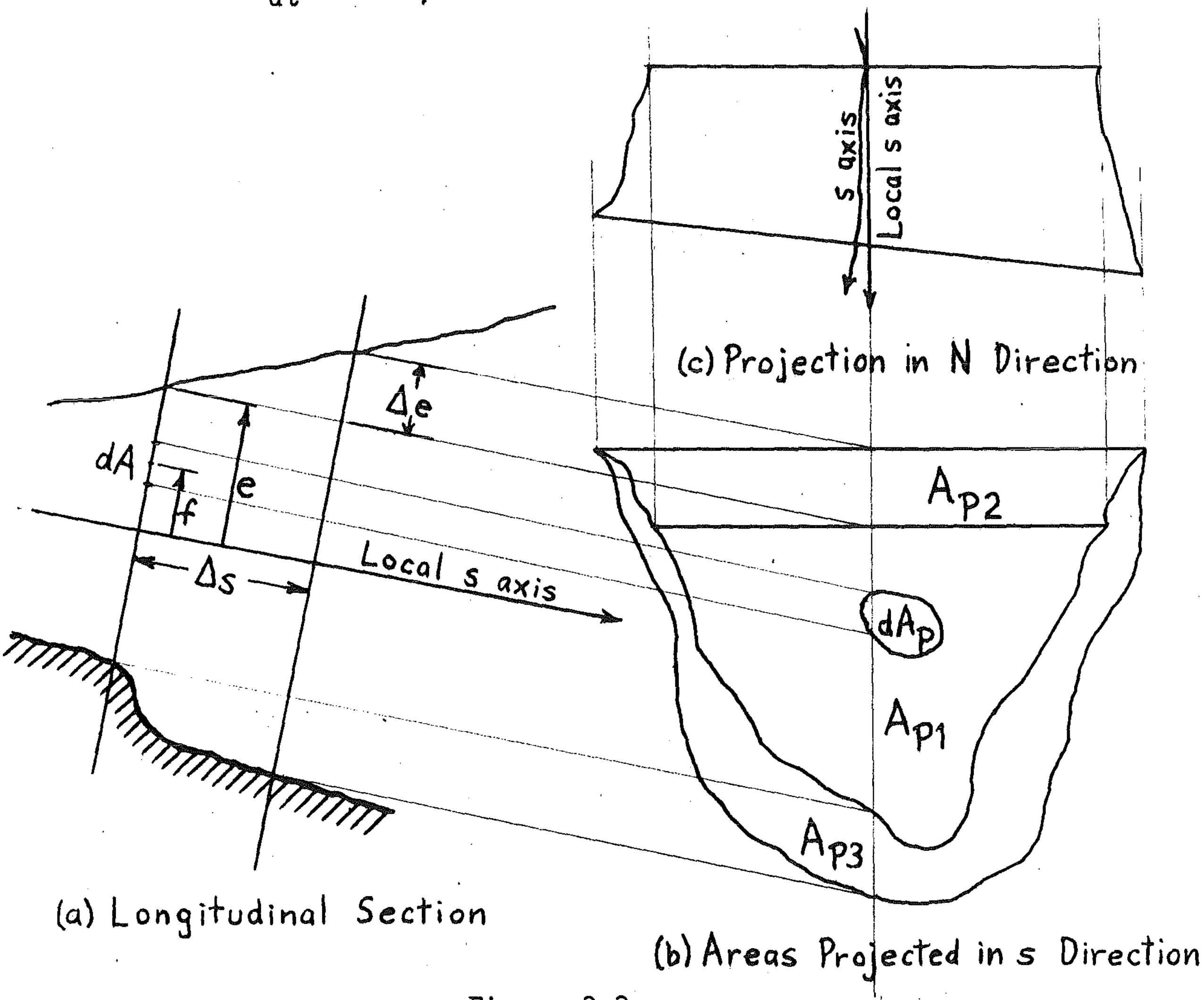



Figure 2-2

Now, after time $\Delta t$, the trailing face of the element is travelling with velocity $v + vv_s\Delta t + v_t\Delta t$,* while the leading face travels with velocity $v + v_s\Delta s + vv_s\Delta t + v_t\Delta t$* + higher order terms at the same time. Therefore, after time $\Delta t$, $\Delta s$ has become $\Delta s + v_s\Delta s\Delta t$. We may therefore write, as $\Delta t \to 0$

$$\frac{d\Delta s}{dt} = v_s\Delta s \tag{2-3}$$

As dA is a function of s and t, we can write

$$\frac{d(dA)}{dt} = \frac{ds}{dt}(dA)_s + (dA)_t$$

and $\frac{ds}{dt} = v$ as we are "following the fluid"

$$\text{i.e.} \quad \frac{d(dA)}{dt} = v(dA)_s + (dA)_t \tag{2-4}$$

Substituting (2-3) and (2-4) in (2-2) and dividing by $\rho\Delta s$, we get

$$(vdA)_s + (dA)_t = p \tag{2-5}$$

To obtain the continuity equation for the length of streamtube shown in Figure 2-2(a) we must add contributions from all small fluid elements, of which the one treated above is typical. Now vdA is the discharge through the area dA, so that $(vdA)_s$ is the rate of change with s of the discharge through each element. When summed this becomes simply $Q_s$. Similarly

---

* Notation: Throughout this thesis, subscript characters t, s, x or y refer to the partial derivative of the variable with respect to that subscript, e.g. $v_t \equiv \frac{\partial v}{\partial t}$. Any other subscript character is used solely as a means of distinction between similar variables.

$(dA)_t$ becomes $A_t$. When the lateral inflows/unit length p are summed all contributions from inside the streamtube cancel between adjacent elements and there remains only the lateral inflow/unit length through the walls of the streamtube which we shall call q.

Thus the continuity equation becomes for the streamtube of Figure 2-1,

$$Q_s + A_t = q \tag{2-6}$$

Equation (2-6) could indeed have been derived directly without appeal to the smaller element, but it is necessary to use the smaller element in the derivation of the equation of motion which follows, if the velocity is not uniform over each cross section. The above derivation was therefore used both for consistency of treatment and because (2-2) is re-used below.

## 2.3 The Equation of Motion

Newton's second law equates the rate of increase of linear momentum of a closed system to the net force component acting on the system. In the case under study, our system is not closed as momentum is added to our element from the external source of lateral inflow, so that separate provision must be made for this additional supply of momentum. In other words, we now equate rate of increase of momentum in an element to forces acting plus rate of external momentum inflow. Notice that, according to our assumption that curvatures of the streamlines are small, we are using the local s axis as an adequate substitute for the s axis in Figure 2-2.

Referring again to the small element of volume $dA\Delta s$, we have, in the direction of the local s axis,

$$\frac{d(\rho dA\Delta sv)}{dt} = F_w \sin\phi + F_R - F_f + \rho p\Delta su \qquad (2\text{-}7)$$

where $F_w$ is the weight force, $F_R$ is the s component of the resultant of the forces normal to the surface of the element, and $F_f$ is the s component of the resultant of the shear forces tangential to the surface of the element expressed as a drag in the opposite direction to the flow.

Now using equation (2-2) and the fact that we are still "following the fluid".

$$\frac{d(\rho dA\Delta sv)}{dt} - \rho p\Delta su = \rho v\frac{d(dA\Delta s)}{dt} + \rho dA\Delta s\frac{dv}{dt} - \rho p\Delta su$$

$$= \rho(v-u)p\Delta s + \rho dA\Delta svv_s + \rho dA\Delta sv_t \qquad (2\text{-}8)$$

To obtain the equation of motion for the larger streamtube we must again sum the individual contributions. $F_w$ sums simply to the total weight $\rho gA\Delta s$, while $F_f$ sums to $F_f^*$, the s component of the bed shear force and any surface shear such as that caused by wind. $F_R$ sums to $F_R^*$, the s component of the resultant of the forces normal to the surface of the length of streamtube of Figure 2-2(a). $F_R^*$ is evaluated by noting that the component of a force resulting from a pressure on an area is equal to the pressure multiplied by the projection of that area on a plane normal to the desired component. Thus, referring to the areas projected parallel to the s axis in Figure 2-2(b), we have, from (2-1), an s component of force $\left[\rho g(e-f)\cos\phi + b\right] dA_P$ contributed by $dA_P$ when $dA_P$ is in $A_{P1}$. b is the barometric pressure, assumed constant. If $dA_P$ is in $A_{P2}$ the force contribution is $b dA_P$. Finally, when $dA_P$ is in $A_{P3}$ the force contribution is $\left[\rho g(e + J\Delta e - f)\cos\phi + b\right] dA_P$, where J is introduced because $dA_P$ may originate from any point along $\Delta s$ as is suggested in Figure 2-2(a). Clearly $0 \leqslant J \leqslant 1$. Therefore

$$
\begin{aligned}
F_R^* = & \int_{A_{P1}} \left[\rho g(e - f)\cos\phi + b\right] dA_P + \int_{A_{P2}} b dA_P \\
& + \int_{A_{P3}} \left[\rho g(e + J\Delta e - f)\cos\phi + b\right] dA_P \\
& - \int_{A_{P1} + A_{P2} + A_{P3}} \left[\rho g(e + \Delta e - f)\cos\phi + b\right] dA_P
\end{aligned}
$$

Note the last term, the force component on the leading face of the element, is evaluated in the same way as that on $A_{P1}$.

Simplifying,

$$
\begin{aligned}
F_R^* = & -\int_{A_{P1}} \rho g \Delta e \cos\phi \, dA_P - \int_{A_{P2}} \left[\rho g(e + \Delta e - f)\cos\phi\right] dA_P \\
& - \int_{A_{P3}} \rho g(1 - J)\Delta e \cos\phi \, dA_P \\
= & -\rho g A_{P1} \Delta e \cos\phi
\end{aligned}
$$

neglecting the last two integrals because $A_{P2}$ and $A_{P3}$ are an order of magnitude smaller than $A_{P1}$, while $e + \Delta e - f$ is less than $\Delta e$ over $A_{P2}$ and $(1 - J)\Delta e$ is less than $\Delta e$ over $A_{P3}$.

Since $A_{P1} = A$, we have

$$F_R^* = -\rho g A \Delta e \cos\phi \tag{2-9}$$

and this holds for more complicated configurations of projected areas than that in Figure 2-2(b) if we treat the sub-areas in a consistent manner.

This result was not unexpected, as it arises directly from the assumption of constant piezometric pressure over a section normal to the s axis, plus the assumption that the streamlines are approximately parallel to the streamtube axis which allows us to conclude that $A_{P2}$ and $A_{P3}$ are an order of magnitude smaller than $A_{P1}$ if $\Delta s$ is small. Indeed an equivalent expression is often produced without any comment as a trivial deduction, but frequently a "bed slope" is introduced which can cloud the simplicity of (2-9). It is therefore considered worthwhile to labour the derivation at this point in order to stress that analysis of the problem is unaffected even if little detailed information is available about the bed geometry.

To evaluate the remaining two terms of (2-7) for the streamtube, equation (2-8) is used. We define U such that the sum of the terms $\rho(v - u)p\Delta s$ from all the elements is $\rho qU\Delta s$.

That is, $$qU = \sum p(v - u) \qquad (2\text{-}10)$$

because $\rho$ and $\Delta s$ are constant for all elements.

The two remaining terms become

$$\rho qU\Delta s + \int_A \rho\Delta s v v_s \, dA + \int_A \rho \Delta s v_t \, dA$$

Thus for the large streamtube, the equation corresponding to (2-7) becomes, after division by $\rho A\Delta s$

$$\frac{qU}{A} + \frac{1}{A}\int_A v v_s \, dA + \frac{1}{A}\int_A v_t \, dA = g \sin\phi - g\frac{\Delta e}{\Delta s}\cos\phi - \frac{F_f^*}{\rho A\Delta s} \qquad (2\text{-}11)$$

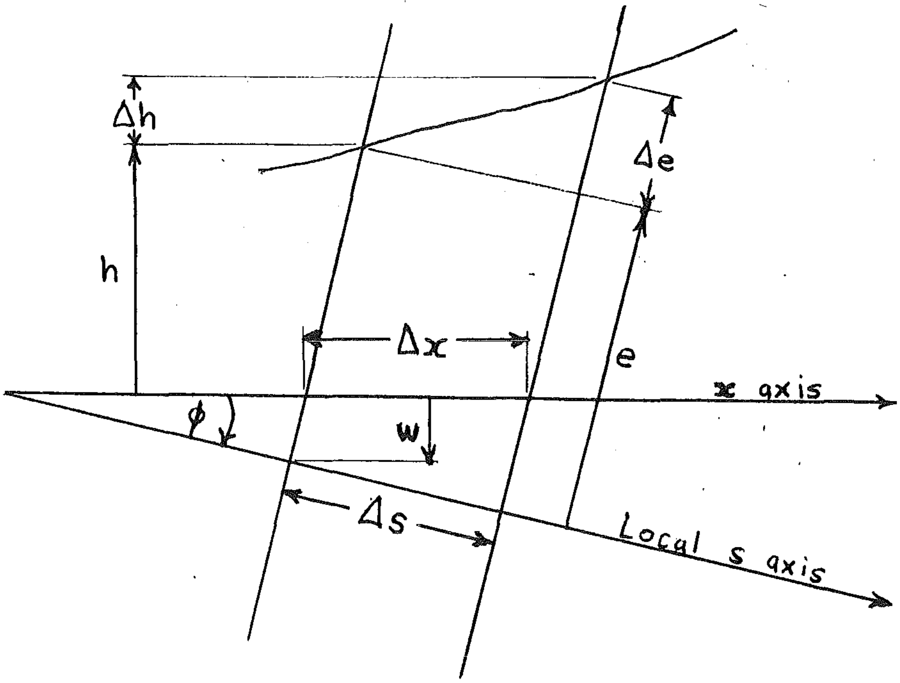



Figure 2-3

We now examine $\Delta e \cos\phi$.

Let $e \cos\phi = h + w$ as shown in Figure 2-3. h is the vertical distance of the surface above the x axis and w is the vertical distance of the s axis, at the cross section considered, below the x axis.

Therefore $dw/ds = \sin\phi$

We see $\Delta e \cos\phi = \Delta h + \Delta s \sin\phi$ so that

$$g \sin\phi - g \frac{\Delta e}{\Delta s} \cos\phi = -g \frac{\Delta h}{\Delta s}$$

$$= -gh_s \qquad (2\text{-}12)$$

as $\Delta s$ tends to zero. The partial derivative is used because $\Delta e$ and hence $\Delta h$ have been defined as existing at the instant at which we are resolving the instantaneous forces on the element.

Because $h_s$ is a slope we introduce the familiar friction slope $S_f$ at this point for consistency of terminology.

$$\text{We define } S_f = \frac{F_f{}^*}{\rho g A \Delta s} \tag{2-13}$$

Note that we have as yet made no effort to evaluate $S_f$ : this is postponed to Section 2.9.

As we normally wish to express equation (2-11) in terms of the mean velocity $V = Q/A$, we set

$$\beta = \frac{\int_A v v_s \, dA}{V V_s A} \tag{2-14}$$

and

$$\eta = \frac{\int_A v_t \, dA}{V_t A} \tag{2-15}$$

Using (2-12) to (2-15) in (2-11) we reach our general equation of motion in the s direction for the streamtube:

$$\frac{qU}{A} + \beta V V_s + \eta V_t = -g h_s - g S_f \tag{2-16}$$

## 2.4 The Definition of the Axes

It is worth clarifying the definition of our system of axes in the light of Section 2.3 before we examine the variables in the two fundamental equations, (2-6) and (2-16).

We have used the "streamtube axis" as our basic direction, relying on an intuitive understanding of this expression. Because we are concerned with the resolution of forces on a fluid element, the most accurate definition of the direction of the streamtube axis is the direction of the resultant of the forces on a large water element. The "large water element" may be defined in the absence of an s axis as any element defined by two sections whose free surface is horizontal across the streamtube. The location of the s axis on the cross section is not important as we are dealing with rates of change (slopes) rather than absolute distances.

Fortunately, the s axis direction is not required with any great accuracy unless $\phi$ is large, as a small error $\delta\phi$ in $\phi$ would mean that we would be applying the force/momentum principle in (2-7) using the components of the true force and momentum vectors along the incorrect s axis. This would obviously make little difference to the vector component magnitudes if $\delta\phi$ were small except where $\cos\phi$ appears in (2-9). The value of $\sin\phi$ in (2-7) would also be affected but these two functions of $\phi$ are combined in (2-12).

Thus only the term $-gh_s$ in (2-16) could be affected markedly by a small error in the direction of the s axis. As $dx/ds = \cos\phi$,

$$h_s = h_x \cos\phi \qquad (2\text{-}17)$$

Now h is measured from the horizontal x axis, so the small error $\delta\phi$ would in effect mean setting

$$h_s = h_x \cos(\phi + \delta\phi)$$

which is not noticeably different from (2-17) unless $\phi$ is large. If $\phi$ is large, errors in $\phi$ would usually be less significant than other errors, so the s axis may be defined adequately for any flow of the type described in Section 2.1.

There is one point in our analysis which needs clarification. Early in Section 2.1 we decided that accelerations normal to the s axis could be ignored if the streamlines had little curvature and if the lateral inflow had negligible velocity normal to the s axis. The second condition does not hold in the case, for instance, of heavy rain falling on an overland sheet flow and obviously a pressure gradient is caused by the deceleration of this inflow to a zero velocity normal to the bed. However, if we treat this pressure gradient as superimposed on the constant piezometric pressure in the N direction, and uniform along the s axis, we are able to suppose that the net contribution of this gradient to our analysis of forces in the s direction may be neglected. A pressure gradient of this type in the H direction is less likely, but might most simply be treated by redefining the H axis parallel to the lines of equal pressure, including the free surface. The N axis would accordingly rotate about the s axis so that the remainder of the analysis would be largely unchanged, provided that the rotation was small.

## 2.5 Dependent Variables in the Overland Flow Equations

We require the Overland Flow equations to be a determinate system of simultaneous equations based on the continuity equation (2-6) and the equation of motion (2-16). Accordingly we seek further relationships between the dependent variables in these equations so that the number of equations equals the number of dependent variables.

In the continuity equation (2-6) we have the variables Q, s, A, t and q, where s and t are the independent variables of the problem. Now Q = AV by definition, so that we can take V, A and q as our dependent variables. q is often unconstrained by the conditions in the streamtube, so we shall assume that q may be specified by some separate relationship between q and s, t and possibly V and A. We denote this relationship by

$$q \quad = \quad q(s,t,V,A) \equiv q(I) \tag{2-18}$$

where this relationship is part of the information needed to solve a given overland flow problem. Thus the continuity equation effectively contains only two dependent variables, V and A. We now turn to the equation of motion (2-16).

## 2.6 Evaluation of the Momentum Terms

The momentum terms (L.H.S.) in equation (2-16) include three new parameters, U, $\beta$ and $\eta$. These parameters are not easy to evaluate except in special cases, so that further assumptions are usually necessary. Some possibilities are now examined.

If we assume $\quad p \quad = \frac{q\,dA}{A}$ (2-19)

and u is constant i.e. the inflow is uniformly distributed with a constant s component of velocity over the whole cross section, then from (2-10)

$$\begin{aligned} qU &= \sum q \,\frac{dA}{A}\,(v - u) \\ &= q\,\left(\sum\frac{v\,dA}{A} - u\sum\frac{dA}{A}\right) \end{aligned}$$

so that $$qU = q(V-u) \tag{2-20}$$

With regard to the variation in velocity across the cross section, it is reasonable to assume a velocity profile such that

$$v = iV \tag{2-21}$$

where i is independent of s and t along any small streamtube but may vary from streamtube to streamtube.

From (2-14) $$\beta = \frac{\int_A iViV_s dA}{VV_s A}$$

$$= \frac{\int_A v^2 dA}{V^2 A} \tag{2-22}$$

which is the definition of the standard momentum coefficient $\beta$, sometimes called the Boussinesq coefficient.

From (2-15) $$\eta = \frac{\int_A iV_t dA}{V_t A}$$

$$= \frac{\int_A v dA}{VA}$$

$$= 1 \tag{2-23}$$

Thus, without necessarily implying that (2-20), (2-22) and (2-23) hold, we may say that

$$U = U(s,t,V,A) \equiv U(I) \tag{2-24}$$

and similarly,

$$\beta = \beta(I) \tag{2-25}$$

$$\eta = \eta(I) \tag{2-26}$$

## 2.7 Evaluation of the term $-gh_s$

The term $-gh_s$ expresses the effect of the forces normal to the surface of the streamtube on the equation of motion, plus the effect of the gravity force on the flow. We must clearly find one relationship between h, A and V to make the problem determinate. If the channel bed is fixed, there is normally at any cross section a one to one relationship between A and the maximum flow depth, and therefore between A and any ordinate -y in the N direction (Figure 2-1), from the water surface to a fixed datum. We therefore seek to establish the datum for any cross section such that the functional relationship between A and y is as simple as possible - for instance "stage" above some arbitrary water level recorder datum is often convenient because measured stage/discharge records of a river may be available. In any case we define

$$z = y \cos\phi - h \tag{2-27}$$

so that z is the vertical distance of the x axis above our datum. Hence

$$-gh_s = gz_s - gy_s \cos\phi \tag{2-28}$$

remembering that we are assuming that the curvature of the s axis is small and so $(\cos\phi)_s$ is negligible.

If the channel bed is fixed the datum slope $z_s$ is a function of s alone, but it may be possible to find

$$z_s = z_s(s,t,V,A) \equiv z_s(I) \tag{2-29}$$

as required even in a channel with a mobile bed. The datum in natural channels can often be taken to be a short fixed distance below the water surface at a low steady flow, so that the datum slope then equals the water surface slope at that low flow. Note that the common term "bed slope" is avoided in this thesis as it is ill defined for natural channels and even for artificial channels which are non-prismatic, except when they are rectangular in cross section.

In a channel with a fixed bed we may write

$$A = A(y(s,t),s) \tag{2-30}$$

Provided (2-30) represents a one to one relationship between A and y, it can be regarded as a special case of the general expression

$$y_s = y_s(s,t,V,A) \equiv y_s(I)$$

It may not be necessary to postulate a fixed bed to obtain such a relationship, but as yet the relationship between V, A, y and z has not been fully resolved (Henderson (1966) Sec. 10.2) for channels with mobile beds.

## 2.8 Simple Channels

Equation (2-30) may still represent relationships which are difficult to manipulate algebraically during the evaluation and solution of overland flow problems. In a channel with a fixed bed there are four geometrical properties of importance in a cross section. These are A and y as already discussed, plus B, the surface width, and P, the wetted perimeter. The influence of P is commonly expressed as a property of the hydraulic radius R, where R is defined by

$$R = \frac{A}{P} \tag{2-31}$$

Note that for a channel with a fixed bed

$$R = R(y,s) \tag{2-32}$$

A prismatic, wide rectangular channel is commonly assumed for simplicity, where "wide" means that $y \ll B$. In this case we have

$$A = By \tag{2-33}$$

$$R = y \tag{2-34}$$

However, we can define k and L such that

$$kA = By \tag{2-35}$$

$$LR = y \tag{2-36}$$

without implying any restriction on k and L, but we also have relationships no more difficult in principle than (2-33) and (2-34) if k and L can be regarded as constants at a cross section. We shall therefore describe channels for which k and L are constant at any cross section as "simple channels". Clearly $B = A_y$, so equation (2-35) means that a simple channel cross section obeys the equation

$$A = Ey^k \tag{2-37}$$

where $E = E(s)$, $k = k(s)$

Simple channel cross sections for a range of k are shown in Figure 2-4.

Note that the cross sections need not be symmetrical, as the subarea on one side of the deepest point may be larger than that on the other side. E has a physical meaning dependent on k - for instance, when $k = 1$, E is the channel width B; when $k = 2$, E is the average side slope of the channel walls.

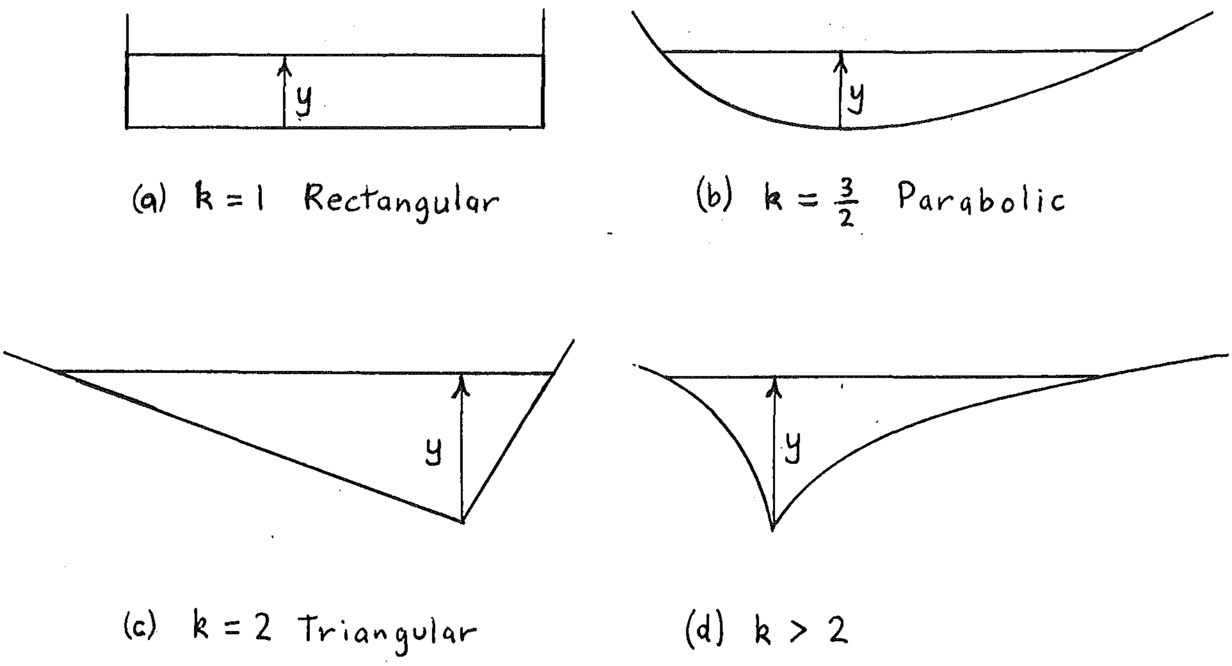



Figure 2-4

By our definition of simple channels, we require the cross sections of Figure (2-4) to be consistent with L constant in (2-36) as well as k constant in (2-35). This is clarified by rearranging (2-36) and (2-35) to give

$$L = \frac{Py}{A} = k\,\frac{P}{B} \tag{2-38}$$

Thus if k is constant, L will also be constant if P/B is constant. This is assured for wide channels where P = B, in which case L = k. However equation (2-38) also gives L constant for all triangular channels, so that theory derived for simple channels is not restricted to wide channels.

It can be seen that many open channels may be approximated by simple channels, and this concept is frequently invoked in the remainder of the thesis, particularly in Section 2.9, Chapter 3, Chapter 5 and Chapter 8.

## 2.9 Evaluation of $S_f$

Empirical formulae have been developed over the last two centuries which express $S_f$ in terms of V and geometrical properties for turbulent flow in open channels. One expression for $S_f$ is the Chézy equation which can be expressed

$$S_f = \frac{V^2}{C^2 R} \tag{2-39}$$

where C is a constant function of channel properties in any reach and R is the hydraulic radius as defined by (2-31). Another common expression for $S_f$ is the Manning formula

$$S_f = \frac{V^2}{C^2 R^{4/3}} \tag{2-40}$$

where $$C = \frac{M}{n} \tag{2-41}$$

and M is a dimensioned constant $= 1.00$ metre$^{\frac{1}{3}}$/sec

$= 1.49$ feet$^{\frac{1}{3}}$/sec

and n is the well known "Manning n", a dimensionless constant related to the roughness of the bed surface in any reach.

While these formulae have generally been validated for uniform steady flows only, there seems no reason why the nature of the surface drag on an element of water should change markedly for the gradually varied unsteady flow which we are considering here. The Manning formula has in fact been tested in the laboratory (Sarma and Sasikanth (1965)) with steady, gradually varied flows over surfaces of different roughness, and the behaviour of the Manning n was consistent with well established (Henderson (1966) p.99) uniform flow results.

For laminar flow, we evaluate $S_f$ analytically under simplified conditions as follows.

Referring again to the small element of volume $dA\Delta s$, we may express (2-7) as

$$F_f = F_w \sin\phi + F_R - \rho(v-u)p\Delta s - \rho dA\Delta s v v_s - \rho dA\Delta s v_t$$

From the argument leading up to (2-9) we may use

$$F_R = -\rho g dA\Delta e \cos\phi$$

and hence, using (2-12) and dividing by $\rho g dA\Delta s$ we get

$$\frac{F_f}{\rho g dA\Delta s} = -h_s - \frac{p(v-u)}{g dA} - \frac{v v_s}{g} - \frac{v_t}{g}$$

and this becomes, assuming a velocity distribution $v = iV$ as in (2-21),

$$\frac{F_f}{\rho g dA\Delta s} = -h_s - \frac{p(iV-u)}{g dA} - \frac{i^2 V V_s}{g} - \frac{i V_t}{g} \tag{2-42}$$

Now any laminar flow to which it will be worth applying the present analysis will be wide and shallow. In such flows each isotach - line of constant velocity on the cross section - can be regarded as being everywhere the same distance, measured in the N direction, from the channel bed. Thus any change of slope in the isotachs will lie directly above the same change of slope in the channel bed, as shown in Figure 2-5(a), so that any isotach can be constructed simply by raising the channel bed, without distortion, in the -N direction. This assumption is of doubtful validity near channel walls or bed irregularities, such as that indicated in Figure 2-5(a), which approach vertical side slopes, as the isotachs would crowd together at these points. However the influence of such features in a "wide" channel is commonly regarded as negligible.

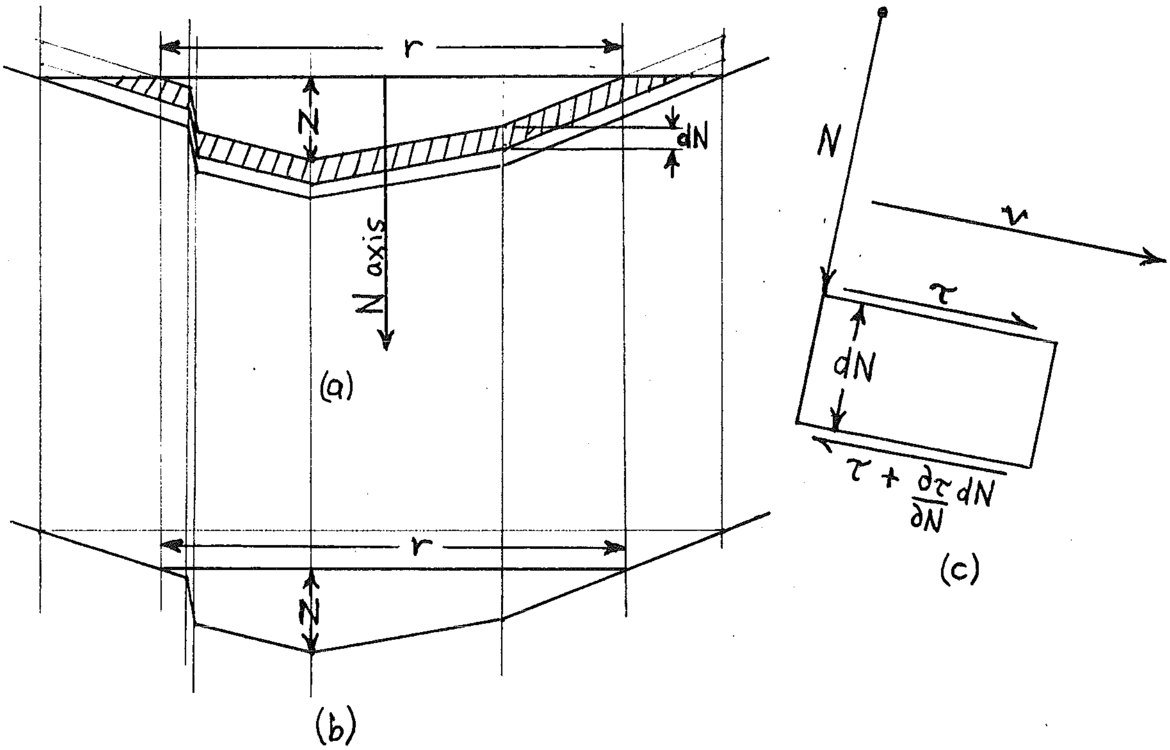



Figure 2-5

Each isotach may be identified by the distance N, along the N axis, of its deepest point below the horizontal free surface. As the area dA in equation (2-42) we can choose that between two isotachs distance dN apart at all points. As this distance is measured only in the N direction, dA = rdN, where r is the horizontal length of the isotach (Figure 2-5(a)). If the bed is fixed r is a function of N alone at any section, provided N is always measured from the free surface. This is true for any y, because any isotach can also be constructed under our assumptions by lowering the free surface in the N direction as shown in Figure 2-5(b). In other words r(N,s) is the free surface width if the water were flowing at maximum depth N in the same channel bed.

Evaluating $F_f$, of equation (2-42), for an element $\Delta s$ in length with dA = rdN as cross-sectional area we have

$$F_f = \frac{\partial \tau}{\partial N} r dN \Delta s$$
$$= -\mu \frac{\partial^2 v}{\partial N^2} r dN \Delta s \qquad (2\text{-}43)$$

since $\tau$, the shear stress on the surfaces of the element acting in the way defined in Figure 2-5(c), is defined for a Newtonian Fluid (Rouse (1950) p.76)

$$\tau = -\mu \frac{\partial v}{\partial N} \qquad (2\text{-}44)$$

$\mu$ is the viscosity and the negative sign arises from the fact that v decreases with increase in N, whereas $\tau$ and $\mu$ are positive. From (2-43)

$$\mu \frac{\partial^2 v}{\partial N^2} = -\rho g \frac{F_f}{\rho g dA \Delta s} \qquad (2\text{-}45)$$

To integrate this, the right hand side must be known as a function of N. We cannot use (2-42) as this is the very equation we wish to solve for V by substituting for $F_f$, and i varies in some as yet undetermined way across the profile. If however in (2-42) we take an average value of i (and $\frac{p(iV-u)}{dA}$) across the section such that $F_f/dA$ may be regarded as constant across the section, we can integrate (2-45). This assumption means, see (2-13), that we assume

$$\frac{F_f^*}{A} = \frac{F_f}{dA} = \rho g S_f \Delta s \tag{2-46}$$

We can express the validity of this assumption relative to a dimensionless number

$$X = \frac{\frac{p(iV-u)}{dA} + i^2 VV_s + iV_t}{gh_s} \tag{2-47}$$

which compares the importance of the momentum terms with that of the pressure and weight forces.

Because $h_s$ is constant over the section, our assumption will be exact if X is zero, for instance in steady uniform flows. As X increases our assumption becomes less accurate, and if X is large we may make our assumption (2-38) only if the numerator of X is reasonably constant over the cross-section.

However, in laminar flow as it is usually understood the momentum terms are indeed small, and therefore it will be understood for "laminar flow" in the rest of this thesis that X is small and therefore (2-46) holds.

Using (2-46) in (2-45), we have

$$\mu \frac{\partial^2 v}{\partial N^2} = -\rho g S_f$$

Integrating, $\mu \frac{\partial v}{\partial N} = -\rho g S_f N + C_1 = -\tau$

Assuming wind shear $\tau = T$ when $N = 0$, $C_1 = -T$

i.e. $$\tau = \rho g S_f N + T \tag{2-48}$$

Integrating again, $\mu v = -\rho g S_f \frac{N^2}{2} - TN + C_2$

$v = 0$ when $N = y$, therefore $C_2 = \rho g S_f \frac{y^2}{2} + Ty$

so that $$\mu v = \frac{\rho g S_f}{2}(y^2 - N^2) + T(y-N) \tag{2-49}$$

Now $$AV = \int_0^y vr\,dN$$

$$= \int_0^y \left[\frac{\rho g S_f}{2\mu}(y^2 - N^2) + \frac{T}{\mu}(y-N)\right] r\,dN \tag{2-50}$$

which establishes $S_f$ as a function of y, A and V provided T is specified over the solution.

For a simple channel (Section 2.8) where $A = Ey^k$,

$$B = A_y = kEy^{k-1}$$

Hence $$r = kEN^{k-1} \tag{2-51}$$

because, as discussed above, r is the free surface width for water of depth N.

Substituting from (2-51) into (2-50) we get for $T = 0$, after integrating,

$$AV = \frac{\rho g S_f}{2\mu}\left(\frac{y^{k+2}}{k} - \frac{y^{k+2}}{k+2}\right)kE = \frac{\rho g S_f E y^{k+2}}{\mu(k+2)}$$

Hence $$S_f = \frac{V\mu(k+2)}{\rho g y^2}$$

$$= \frac{V\mu(k+2)}{\rho g k^2 R^2} \tag{2-52}$$

because for a wide simple channel $y = kR$, from (2-38).

The expressions for $S_f$ (2-39), (2-40) and (2-52) are all included in the general expression

$$S_f = \left(\frac{V}{DR^m}\right)^{1/j} \qquad (2\text{-}53)$$

$$\text{or} \qquad V = DR^m S_f{}^j \qquad (2\text{-}54)$$

where j, D and m are constants. Note that $S_f$ always takes the same sign as V, where V is taken positive in the direction of increasing s.

We therefore take it that $S_f = S_f(s,t,V,R) \equiv S_f(I)$ (2-55)

## 2.10 The Overland Flow Equations

We are now in a position to incorporate the continuity equation (2-6) and the equation of motion (2-16) into a determinate system of simultaneous equations, the Overland Flow equations. These are:

From (2-6) and (2-18)

$$VA_s + AV_s + A_t = q(I) \qquad (2\text{-}56)$$

From (2-16), (2-18), (2-24), (2-25), (2-26), (2-28), (2-29) and (2-55)

$$\frac{q(I)U(I)}{A} + \beta(I)VV_s + \eta(I)V_t = gz_s(I) - gy_s\cos\phi - gS_f(I) \qquad (2\text{-}57)$$

Plus (2-30) and (2-32)

$$A = A(y(s,t),s) \qquad (2\text{-}30)$$

$$R = R(y(s,t),s) \qquad (2\text{-}32)$$

Thus we have four equations in four dependent variables V,A,y and R, because we have used the notation f = f(I) to describe a functional relationship with some or all of s, t, V and A (or R, or y), which must be specified independently as part of the information needed to solve the problem. We have retained the four dependent variables to this point in order to stress that equations (2-56) and (2-57) are not necessarily dependent on the assumption of a fixed channel bed although (2-30) and (2-32) were introduced on the basis of that assumption. However, because the relationships between A, y and R, and the evaluation of $z_s$, are obscure at present for a channel with a mobile bed, we shall hereafter take it that the channel bed does not move significantly, so that the Overland Flow equations become essentially a determinate system of the equations (2-56) and (2-57) in two dependent variables V and y.

# PART II - FINITE DIFFERENCE SOLUTIONS

## CHAPTER 3

## EXISTING SOLUTION THEORY

### 3.1 Rearrangement of the Overland Flow Equations

We now rearrange the Overland Flow Equations (2-56) and (2-57) into a convenient form for solution. From (2-30) and the fact that the surface width $B = A_y$,

$$A_t = By_t \tag{3-1}$$

We shall define

$$H_s = A_s - By_s \tag{3-2}$$

so that $H_s$ is the variation of area with s for y constant. B and $H_s$ are in general functions of y and s, which may be expressed using a polynomial representation of A in place of (2-30).

$$A = E_0 + E_1 y + E_2 y^2 + E_3 y^3 + \dots \tag{3-3}$$

where $E_0$, $E_1$, $E_2$ .... are functions of s alone and depend on the shape of the fixed channel bed. We may clearly express any one to one A-y relationship at all cross sections by such a polynomial. Hence in this case

$$B = E_1 + 2E_2 y + 3E_3 y^2 + \dots \tag{3-4}$$

$$H_s = (E_0)_s + y(E_1)_s + y^2(E_2)_s + \dots \tag{3-5}$$

We use (3-1) and (3-2) in (2-56) and (2-57) to give

$$VBy_s + By_t + AV_s = q - VH_s \tag{3-6}$$

$$g \cos \phi \; y_s + \beta VV_s + \gamma V_t = gz_s - gS_f - \frac{qU}{A} \tag{3-7}$$

At any time instant we can assess every quantity except $y_t$ and $V_t$ directly if we know y and V for all s. Thus we can solve for $y_t$ from (3-6) and for $V_t$ from (3-7). This fact has been used in elementary applications of the method of finite differences, where $y_t$ and $V_t$ have been replaced by the forward differences $\frac{y(s,t + \Delta t) - y(s,t)}{\Delta t}$ and $\frac{V(s,t + \Delta t) - V(s,t)}{\Delta t}$ respectively. Hence the solution is apparently advanced to time $t + \Delta t$ and the process can be repeated. Unfortunately this approach relies on the unsupportable assumption that the values of $y_t$ and $V_t$ at the instant t = t persist for long enough to be treated as averages over the succeeding time step $\Delta t$, whereas in fact $y_t$ and $V_t$ are subject to discontinuities, for instance if the lateral inflow q varies with time in a discontinuous manner.

Even if q varies in a continuous manner, discontinuities in the derivatives may propagate through the flow. This arises from the fundamental nature of the Overland Flow equations, as we now show.

## 3.2 Characteristics in the Overland Flow Equations

Steady flow at a cross section of a channel will be disturbed if the flow at the upstream end of the channel is altered. However, if that section is some distance from the upstream end, a significant period of time elapses from the introduction of the disturbance before any effect is detectable on the flow at the section. In other words a disturbance propagates its influence at some finite speed down the channel. This is the essence of the theory of characteristics, that in a physical system which is subject to delayed influence from externally imposed changes at the system boundaries there exist paths along which such influence propagates at a finite speed related to the properties of the system. These paths are called characteristics.

Mathematically, the influence of disturbances must propagate at velocity ds/dt along the characteristics crossing the s-t plane. Using the definition of partial differentiation, we have

$$ds V_s + dt V_t = dV \tag{3-8}$$

$$ds y_s + dt y_t = dy \tag{3-9}$$

Thus (3-6), (3-7), (3-8) and (3-9) form a system of four equations which may be written in matrix form:

$$\begin{bmatrix} VB & B & A & 0 \\ g\cos\phi & 0 & \beta V & \eta \\ 0 & 0 & ds & dt \\ ds & dt & 0 & 0 \end{bmatrix} \begin{bmatrix} y_s \\ y_t \\ V_s \\ V_t \end{bmatrix} = \begin{bmatrix} q - VH_s \\ gz_s - gS_f - \frac{qU}{A} \\ dV \\ dy \end{bmatrix} \tag{3-10}$$

Now we imagine an observer with access to instruments which measure instantaneous values of y and V at a number of sections along a channel. He takes one spot measurement $y_1$ and $V_1$, then a time dt later he measures $y_2$ and $V_2$ using instruments a distance ds along the channel from the section used for the first measurement. The difference $V_2 - V_1$ is $dV$ and $y_2 - y_1$ is $dy$. If ds and dt are small V and y can be regarded as constant, as our assumptions during the derivation of the Overland Flow equations in Chapter 2 implied that discontinuities in V and y are not permitted in the system. Thus the observer can obtain all parameters in (3-10) except $y_s$, $y_t$, $V_s$ and $V_t$. Thus we have four linear equations in four unknowns, a determinate system unless the "rank" (Aitken (1956) p.60) of the matrix of coefficients of the four unknowns is less than four. That is, the system is indeterminate if

$$\begin{vmatrix} VB & B & A & 0 \\ g\cos\phi & 0 & \beta V & \eta \\ 0 & 0 & ds & dt \\ ds & dt & 0 & 0 \end{vmatrix} = 0$$

That is, if

$$\frac{ds}{dt} = \frac{V(\beta + \eta) \pm \sqrt{V^2(\beta - \eta)^2 + 4g\cos\phi\,\frac{\eta A}{B}}}{2\eta} \qquad (3\text{-}11)$$

Now $\eta$ is positive, so that when the ratio of ds to dt takes the two real values given by (3-11), the system (3-10) is no longer determinate. Thus if our observer chooses ds and dt in a ratio satisfying (3-11) he cannot determine the partial derivatives $y_s$, $y_t$, $V_s$, and $V_t$. Therefore if he makes his observations so that each ds and dt form a ratio satisfying (3-11) he is transferring his attention at a speed ds/dt along the channel while the partial derivatives remain indeterminate.

When (3-11) holds, we can test (3-10) for the existence of a solution by replacing any column of the matrix of coefficients by the R.H.S. column vector and requiring the resulting matrix to be singular (see Aitken (1956) p.70). That is, we require

$$\begin{vmatrix} VB & B & q - VH_s & 0 \\ g\cos\phi & 0 & gz_s - gS_f - \frac{qU}{A} & \eta \\ 0 & 0 & dV & dt \\ ds & dt & dy & 0 \end{vmatrix} = 0 \qquad (3\text{-}12)$$

Dividing (3-12) by B $dt^2$ and rearranging,

$$\left[\eta\frac{dV}{dt} - (gz_s - gS_f - \frac{qU}{A})\right](\frac{ds}{dt} - V) + \left[\frac{dy}{dt} - (\frac{q}{B} - \frac{VH_s}{B})\right]g\cos\phi = 0 \qquad (3\text{-}13)$$

where ds/dt is defined by (3-11).

If we replace any other column by the R.H.S. vector column we get an equation equivalent to (3-13) as long as (3-11) holds.

Because (3-11) has two distinct solutions, (3-13) in fact represents two ordinary differential equations (O.D.E's) which must be satisfied along distinct curves on the s-t plane defined by the two roots of equation (3-11). Thus if we can specify initial values of V and y on each of two intersecting curves, V and y must also be defined at the intersection. Conversely, the value of V and y at any section of a channel at any time depends only on the initial values of V and y specified for the curves defined by (3-11) which intersect at that point. Thus the influence of any disturbance in V and y (within the limits set in Chapter 2) travels along those curves, so that (3-11) defines the characteristic curves described earlier.

It is significant that the partial derivatives are indeterminate along these curves, as discontinuities in these derivatives become a possibility along the characteristics, so that the solution cannot be assumed to be analytic. Because such discontinuities do appear with any sudden change in the flow conditions, whether variations in channel width, lateral inflow, or upstream inflow, they are commonly present in overland flow problems, invalidating the elementary finite difference approach discussed in Section 3.1 which took for granted the analyticity of the solution.

### 3.3 Characteristic Forms of the Overland Flow Equations

We now define

$$c = +\sqrt{\frac{g \cos \phi \; A}{\eta B}} \tag{3-14}$$

so that (3-11) becomes

$$\frac{ds}{dt} = \frac{(\beta + \eta)}{2\eta} V \pm \sqrt{V^2 \frac{(\beta - \eta)^2}{4\eta^2} + c^2} \tag{3-15}$$

Now let

$$G = \frac{\beta - \eta}{2\eta} \frac{V}{c} \tag{3-16}$$

We wish to evaluate (3-13) in which the factor $\frac{ds}{dt} - V$ appears. From (3-15) and (3-16)

$$\frac{ds}{dt} - V = c(G \pm \sqrt{G^2 + 1})$$

$$= \pm c \qquad (3\text{-}17)$$

if $G \ll 1$. In this case c is clearly a general form of the standard expression for the propagation speed (relative to the fluid) of a low amplitude surface disturbance in shallow water (Henderson (1966) p.327), as from (3-17) the characteristic velocities are $\pm c$ relative to the fluid. A convenient term for this propagation speed is the "celerity" (Ven te Chow (1959) p.538).

Define the Froude Number

$$F = \frac{V}{c} \qquad (3\text{-}18)$$

We have yet to prove that $G \ll 1$ to establish (3-17).

F will only be large (say greater than 1) in highly turbulent flows in which $\beta-\eta$ is negligible. $\beta-\eta$ is still small for turbulent flows of lower Froude Number, so for all turbulent flows $G \ll 1$.

For laminar flow, we take it that (2-21) holds, so that from (2-22)

$$\beta = \frac{\int_0^y \left[\frac{\rho g S_f}{2\mu}(y^2 - N^2)\right]^2 kEN^{k-1}dN}{\left[\frac{\rho g S_f\, y^2}{\mu(k+2)}\right]^2 Ey^k}$$

$$= \frac{2(k+2)}{k+4} \qquad (3\text{-}19)$$

for a simple channel. We have used (2-49) with $T = 0$, (2-51) and (2-52) as well as the simple channel equation (2-37). From (2-23) $\eta = 1$. so that from (3-16)

$$G = \frac{k}{2(k+4)} F \tag{3-20}$$

for laminar flow in a simple channel. In Chapter 5 we show that laminar flow becomes unstable if $F > k/2$, invalidating the assumptions of Chapter 2, so that for our analysis, using (3-20),

$$G < \frac{k^2}{4(k+4)} \tag{3-21}$$

In the most likely simple channel for laminar flow, the rectangular channel, $k = 1$ so that $G < 1/20$, which justifies our assumption $G \ll 1$. We have therefore established (3-17) for all common flows.

Substituting (3-17) into (3-13), dividing by $\pm c\eta$, and rearranging gives

$$\frac{dV}{dt} \pm \frac{g \cos\phi}{\eta} \sqrt{\frac{\eta B}{g \cos\phi \ A}} \frac{dy}{dt} = \frac{gz_s}{\eta} - \frac{gS_f}{\eta} - \frac{qU}{A\eta} \pm (q - VH_s) \frac{g \cos\phi}{B\eta} \sqrt{\frac{\eta B}{g \cos\phi \ A}}$$

Now if we are to integrate these differential equations conveniently, they should be transformed into an O.D.E. in a single variable along each characteristic. We use a device first introduced by Escoffier and Boyd (1962) and define the "stage variable" w such that

$$\frac{dw}{dt} = \sqrt{\frac{gB \cos\phi}{\eta A}} \frac{dy}{dt} \tag{3-22}$$

That is

$$w_t + \frac{ds}{dt} w_s = \sqrt{\frac{gB\cos\phi}{\eta A}}\, y_t + \frac{ds}{dt}\sqrt{\frac{gB\cos\phi}{\eta A}}\, y_s$$

As (3-22) must hold for all ds/dt, this implies

$$w_t = \sqrt{\frac{gB\cos\phi}{\eta A}}\, y_t \qquad (3\text{-}22a)$$

$$w_s = \sqrt{\frac{gB\cos\phi}{\eta A}}\, y_s \qquad (3\text{-}22b)$$

Equation (3-22) also implies that w is a function of y alone, because dw is zero if dy is zero. Thus the equivalent of the definition of w given by Escoffier and Boyd

$$w = \int_0^y \sqrt{\frac{gB\cos\phi}{\eta A}}\, dy \qquad (3\text{-}23)$$

satisfies (3-22) only if A and B are functions of y which do not vary along the channel i.e. if the channel is prismatic. We cannot therefore use (3-23) to define w in non-prismatic channels as this allows w to vary with s even if y is constant, which violates (3-22). However, as discussed further in Chapter 5, we are concerned with variations in w rather than w itself, so that w need not be defined other than by (3-22). It follows that our Overland Flow equations can be written

$$\frac{d}{dt}(V \pm w) = \frac{gz_s}{\eta} - \frac{gS_f}{\eta} - \frac{qU}{A\eta} \pm (q - VH_s)\frac{c}{A} \qquad (3\text{-}24)$$

or, using (3-17)

$$V_t + w_t + (V + c)(V_s + w_s) = \frac{gz_s}{\eta} - \frac{gS_f}{\eta} - \frac{qU}{A\eta} + (q - VH_s)\frac{c}{A} \qquad (3\text{-}25)$$

$$V_t - w_t + (V - c)(V_s - w_s) = \frac{gz_s}{\eta} - \frac{gS_f}{\eta} - \frac{qU}{A\eta} - (q - VH_s)\frac{c}{A} \qquad (3\text{-}26)$$

(3-25) applies along the characteristic on which $\frac{ds}{dt} = V + c$, while (3-26) applies along the characteristic on which $\frac{ds}{dt} = V - c$, and the two equations together are the characteristic form of the Overland Flow equations. Note that these equations are almost as general as the Overland Flow equations (3-6) and (3-7) as we have introduced only one additional assumption, that $G \ll 1$. We have unfortunately introduced two more important variables connected with the cross section, c and w, which may be tedious to evaluate, but (3-25) and (3-26) enable disturbances to be followed along their propagation paths and are therefore a sounder basis for numerical solution than are (3-6) and (3-7).

In a prismatic simple channel, c and w become straightforward functions of y. From (3-14) and (2-35)

$$c = \sqrt{\frac{g \cos\phi \; y}{\eta k}} \tag{3-27}$$

and from (3-23) and (2-35)

$$\begin{aligned} w &= \int_0^y \sqrt{\frac{g \cos\phi \; k}{\eta}} \, y^{-\frac{1}{2}} dy \\ &= 2kc \end{aligned} \tag{3-28}$$

Thus if a channel can be approximated by a prismatic simple channel, great savings in computations result from the fact that all the geometrical variables in the cross section can be carried through a numerical solution as simple functions of c.

## 3.4 Solution Theory of Hyperbolic Equations

Because the system of Overland Flow equations possesses distinct real characteristics, it belongs to the hyperbolic class of systems of partial differential equations for which a considerable body of solution theory has been developed. As the work in the next three chapters is an extension of this theory a brief account of its main relevant features is now presented.

This solution theory can be divided into two broad categories, the theory of characteristics and the theory of difference methods. The theory of characteristics is discussed thoroughly by Courant and Friedrichs (1967), and applied to long waves in shallow water by Stoker (1957), Chapter 10, and to problems in hydraulics by Henderson (1966), Chapter 8. The theory of difference methods for hyperbolic problems has largely been developed recently, stimulated by the rapid development of computer technology which has made numerical solutions practical instead of merely feasible. Considerable use is made in this thesis of the methods developed by Richtmyer (1957) (see also Appendix A), while Smith (1965) provides a straightforward general survey of the field of numerical analysis.

Although the theory of characteristics provides a valuable means of solving various simplified problems, we are principally concerned here with the insights it offers into the behaviour of the solution of the Overland Flow equations. Such insights are vital to the success of any solution, numerical or otherwise. Now characteristics are paths along which disturbances can propagate, and this is connected with the concepts of "domain of dependence" and "range of influence", which may be defined as follows (see Courant and Friedrichs (1967) Section 24). The initial value problem in this case consists of finding the solution of the Overland Flow equations at a short time after a given initial time $t = t_o$, given a solution along the initial line $t = t_o$ in the s-t plane. By using the concept of ordinary differential equations along the characteristics, it can be shown that the solution at any point P on the s-t plane near the initial line depends only on the initial solution along the segment of the initial line cut off by the two characteristics passing through the point P. This segment is called the domain of dependence. The range of influence of a point P' on the initial line is the totality of points P on the s-t plane which are influenced by the initial data at P'. Any point P which contains P' in its domain of dependence is influenced by the data at P', and hence the range of influence of P' is the region between the two characteristics drawn through P'.

It is the existence of domains of dependence and ranges of influence in wave propagation systems which distinguishes them from other systems, as solutions need not even be analytic. Thus the methods of solving wave propagation systems, whether numerical or otherwise, must be studied from a viewpoint which is fundamentally different from those applying to solutions of ordinary differential equations, or to solutions which are states of equilibrium.

We discussed the possibility of discontinuities* arising from the boundary conditions or from discontinuities in q or channel configuration in Section 3.2. We now see that conditions at a point P whose domain of dependence does not include such a discontinuity cannot be affected by that discontinuity, but if we allow P to move through the solution domain, conditions at P will be affected by the discontinuity immediately P crosses a characteristic which has passed through a discontinuity, as that discontinuity has entered the domain of dependence of P. Thus it may be proved that discontinuities propagate along characteristics, and further that such discontinuities never disappear. This is in conformity with our previous comments on characteristics as lines along which the partial derivatives are indeterminate.

The theory of difference methods for Hyperbolic Partial Differential Equations is thus connected with the theory of characteristics. Characteristics are used to prove the

---

* It should be remembered that our assumption during the derivation of the Overland Flow equations in Chapter 2 implies that discontinuities in depth and velocity are not permitted in the system. Thus any discontinuities, when referred to as such, are always discontinuities of derivatives, which are not excluded by the analysis of Chapter 2.

existence and uniqueness of a solution and are a necessary pre-requisite to the study of the behaviour of a finite difference solution.

Our Overland Flow equations are technically classified as an hyperbolic non-homogeneous system of two quasi-linear first order partial differential equations in two independent variables. The system is "hyperbolic" because we are able to find two real characteristic directions as is shown by equation (3-11), and "non-homogeneous" because terms which do not include derivatives are present. Each equation is of the "first order" because no variable is differentiated more than once, and "quasi-linear" because the coefficients of the first order derivatives are functions of the independent and dependent variables, but not functions of derivatives of the dependent variables.* The two independent variables are of course s and t.

### 3.5 The "Courant Condition"

Courant and Lax (1949) proved for this class of equations the existence and uniqueness of a solution in a "suitably small" region adjacent to a line along which initial conditions were known. This proof used the theory of characteristics. Lax (1953) extended this existence proof and also proved that solutions exist for nonanalytic initial value problems. This confirms our

*Thus because of the continuous nature of the depth and velocity, each may be regarded as "locally constant" in a small region of the solution (see also Appendix A) and hence the equations may be treated in some respects as linear.

expectation that an unique solution must exist over the s-t plane if the Overland Flow equations successfully model the physical flow in which unique (but not necessarily stable) responses undoubtedly exist to any imposed initial and boundary conditions.

Courant, Isaacson and Rees (1952) used the work of Courant and Lax (1949) to prove that finite difference solutions of hyperbolic differential equations can be defined in a suitable small neighbourhood independent of the step size between grid points. These solutions were proved to differ from the exact solution by a bounded quantity of the same order as the step size. It was shown that the finite difference solutions could be based on curvilinear or rectangular lattices of points, but that the finite difference schemes were required to satisfy a criterion, since named the "Courant Condition" e.g. in Liggett and Woolhiser (1967) p.54. We quote this criterion because of its fundamental importance, and also because it illustrates the importance of the theory of characteristics to the understanding of the behaviour of numerical solutions of the Overland Flow equations. The criterion was stated in Section 6, Courant, Isaacson and Rees (1952) as follows:

".... the mesh width ratio .... should be chosen in such a way that the domain of dependence of any point in the mesh as given by the difference equations, is not less than the domain of dependence determined by the differential equations. .... the choice of difference quotients (forward or backward) should be made with the idea of preserving the domain of dependence".

Unfortunately, Courant, Isaacson and Rees did not define the "suitable small neighbourhood" in quantitative terms, although they did show (p.252) that it is restricted in a "timelike" direction (in our case the time direction) to a magnitude inversely proportional to a large positive constant $\gamma$, which is related to the bound of the errors. Thus as our time step $\Delta t$ decreases, the errors and hence $\gamma$ must decrease and hence the "neighbourhood" may expand in time so that the "neighbourhood" certainly contains $\Delta t$ as $\Delta t$ tends to zero. If we increase $\Delta t$ however, $\gamma$ certainly cannot decrease so that at some limiting magnitude $\Delta t$ will pass outside the suitable small neighbourhood in which the proof is rigorous. Therefore the proof does not hold for $\Delta t$ large. Thus there is a restriction on $\Delta t$ implied in this proof, in addition to the Courant condition on $r = \Delta t/\Delta s$, in that $\Delta t$ must be "sufficiently small". We explore a method of establishing a quantitative upper limit on $\Delta t$ in the next three chapters.

Courant, Isaacson and Rees also concluded that the round off error should be of the order $(\Delta t)^2$. That is, if $\Delta t$ is reduced by the ratio $1/d$, the number of decimal places should be increased by $2 \log_{10} d$ digits.

### 3.6 Definitions of Consistency, Convergence and Stability

We introduce the terms consistency, convergence and stability, following Richtmyer (1957).

Consistency:

The consistency condition requires that the difference between the partial differential equations and their finite difference representations should tend towards zero as the mesh lengths are allowed to tend towards zero. Consistency can in practice be investigated by Taylor series expansion of the dependent variables with remainder. For instance, we might wish to represent the simple equation $V_t = \text{const}$

by $$\frac{V_a^{b+1} - V_a^b}{\Delta t} = \text{const}$$

where the notation is as introduced in Appendix A. Taylor series expansion gives

$$V_t - \frac{(V_a^{b+1} - V_a^b)}{\Delta t} = V_t - \frac{(V_a^b + \Delta t V_t + \frac{\Delta t^2}{2} V_{tt} + O(\Delta t^3) - V_a^b)}{\Delta t}.$$

$$= O(\Delta t)$$

which tends to zero as $\Delta t$ tends to zero, so that this finite difference approximation satisfies a consistency condition. We might alternatively represent the same equation by

$$\frac{V_a^{b+1} - \frac{1}{2}(V_{a-1}^b + V_{a+1}^b)}{\Delta t} = \text{const}$$

In this case the same argument gives

$$V_t - \frac{V_a^{b+1} - \frac{1}{2}(V_{a-1}^b + V_{a+1}^b)}{\Delta t} = \frac{(\Delta s)^2}{2\Delta t} V_{ss} + O(\Delta t) + \frac{O(\Delta s)^4}{\Delta t}$$

which does not automatically tend to zero as $\Delta s$ and $\Delta t$ tend to zero, but depends on the relationship between $\Delta s$ and $\Delta t$ in the limit. If the ratio $\Delta s/\Delta t$ remains constant as $\Delta s$ and $\Delta t$ go to zero, the consistency condition is satisfied, but if $(\Delta s)^2/\Delta t$ remains constant the consistency condition is violated.

Thus the second approximation is suspect.

Now for simplicity we have used an equation with a single partial derivative term, and it should be borne in mind that consistency deals with the formulation of finite difference equations rather than individual terms. Thus an equation with many terms may be tested for consistency term by term only if the possibility of compensating single term inconsistencies can be discarded.

The consistency of a number of finite difference schemes for solving the shallow water equations, a special case of the Overland Flow equations, is investigated by Taylor series expansion by Liggett and Woolhiser (1967). Note that their term "approximation" conforms to our definition of "consistency".

Convergence:

The repeated application of a finite difference scheme to a given set of initial and boundary values provides a convergent estimate of the exact solution of the corresponding wave propagation problem if the difference between the finite difference solution and the exact solution tends to zero as the time step $\Delta t$ is allowed to approach zero. This definition is of course required to hold only for the time T over which the solution is carried. We also take it that $\Delta s$ is regarded as a function of $\Delta t$ such that $\Delta s$ tends to zero with $\Delta t$.

Stability:

Unlike convergence, stability refers only to the finite difference solution, and a finite difference scheme is said to be stable if its repeated application cannot lead to unbounded results. In particular, for stability the numerical solution must remain bounded over the solution time T as the mesh lengths $\Delta t$ and $\Delta s$ are allowed to approach zero. Note that it is sufficient under this definition to prove that such a bound exists for the stability condition to be satisfied even though the bound may be very large.

Ideally we should aim to establish convergence and then use mesh lengths approaching zero, and hence our solution would be as exact as we wished. Unfortunately convergence is difficult to investigate, involving as it does the unknown exact solution of the partial differential equations. However Richtmyer (1957) in Chapter III presents Lax's equivalence theorem, which is stated "Given a properly posed initial-value problem and a finite-difference approximation to it that satisfies the consistency condition, stability is the necessary and sufficient condition for convergence". It should be noted that this theorem is proved only for linear initial value problems. Richtmyer, in Chapter IV of his book, then investigates a particular class of linear initial value problems which are amenable to Fourier series methods, and derives the von Neumann necessary condition for stability. Fourier series stability analysis and the von Neumann stability condition are discussed more fully in Appendix A.

For a certain class of linear initial value problems we may therefore investigate convergence by testing for consistency by Taylor series expansion and for stability by Fourier series methods. If the finite difference scheme satisfies both consistency and stability conditions, then convergence of the numerical solution to the exact solution is established for that class of problems by Lax's equivalence theorem. Now our quasi-linear Overland Flow equations may be treated as "locally linear" (see Appendix A) and hence an heuristic application of consistency and stability tests to small areas of the solution should be illuminating. As an example of the application of Fourier series stability analysis to quasi-linear problems, we adapt an analysis given by Richtmyer in Chapter X of his book by analysing the stability of a finite difference scheme applied to Overland Flow equations. This example is presented in Appendix A. This stability analysis gives the same "Courant Condition" as was deduced rigorously by Courant, Isaacson and Rees (1952) for the same finite difference formulation applied to a general quasi-linear system. Thus we conclude that Richtmyer stability analysis is a useful tool in investigating the convergence of numerical solutions to the exact solution of the Overland Flow equations, at least as $\Delta t$ and $\Delta s$ tend to zero.

# CHAPTER 4

## INTRODUCTION OF EXTENDED STABILITY THEORY

### 4.1 Difficulties with "Theoretically Stable" Solutions

There has been a tendency to regard a finite difference solution of an overland flow problem as "theoretically stable" if the mesh ratio $r = \Delta t/\Delta s$ can be shown by Fourier series analysis to satisfy the von Neumann necessary condition for stability (A-14) as $\Delta t$ and $\Delta s$ tend to zero. Such approximate analysis in many cases results in the Courant condition, which was derived rigorously, and encourages the belief that such stability and convergence properties depend only on r even if $\Delta t$ and $\Delta s$ are appreciable. However, as we have seen in Section 3.5, the convergence proof by Courant, Isaacson and Rees (1952) does not hold for $\Delta t$ large. Similarly Richtmyer's methods deal with the stability of the finite difference scheme and its consistency with the partial differential equations as $\Delta t$ tends to zero, but not with the behaviour of the numerical solution for finite values of $\Delta t$. Now in practice we must use a finite value of $\Delta t$ in a computed numerical solution, so that if the restriction on the magnitude of $\Delta t$ is infringed we have no guarantee of convergence.

This would be of little importance if values of $\Delta t$ in common use were found to give satisfactory results with "theoretically stable" finite difference schemes, but this is far from the case.

Stoker (1957) p.492 found in his work on the Ohio River that fluctuations developed in his numerical solution which necessitated small interval sizes being used. It is noticeable that he was able to use a value of r (0.034 hr/mile) in his Region V larger than that (0.024 hr/mile) used in Region II even though the value of V + c was about 50% greater in Region V than in Region II. This is the opposite of the behaviour expected from r if it was governed solely by the Courant condition. Stoker also drew attention to the problem of assessing the maximum value of V + c prior to the solution of a problem, remarking (p.493) that "It would be convenient to be in possession of a safe estimate for the maximum value of the particle velocity, in order to predict an appropriate safe value for the time interval $\Delta t$". We shall discuss the problem of the prior estimation of the particle velocity V + c in Section 4.9, after indicating its importance in the extended stability analysis introduced in this chapter.

Fenzl (1965) found that objectionable oscillations occurred in some numerical solutions where $\Delta t$ was chosen to just satisfy the Courant condition, and suggested that there was a need for a restriction on $\Delta t$ in addition to the Courant condition.

Liggett and Woolhiser (1967), in a paper comparing various finite difference schemes, remarked that for nonlinear equations precise and rigorous stability criteria could not be found, and pointed out that most stability criteria applied only as the mesh lengths tended to zero. They attempted to find stability criteria by experimental means, and concluded that rectangular finite difference schemes are often unsatisfactory even if "theoretically stable". It is of interest that they found that the "Unstable Method", which violates the von Neumann condition, was empirically stable in a few experiments, notably on the rising limb of a runoff hydrograph.

### 4.2 Simple Illustrations of Instability

It is not even necessary to experiment with solutions of the full Overland Flow equations to demonstrate instability for finite $\Delta t$ in solutions using "theoretically stable" difference schemes. We illustrate this by applying the CIR rectangular scheme (see Appendix A) to special cases of overland flows. It is unlikely that such a scheme would be used to solve these simple problems, but an adequate stability theory should predict the possible instabilities which are indicated. We show later that our extended stability theory does in fact predict these instabilities.

We consider first a rectangular tank containing a stationary body of water. Through the two side walls of the tank a constant lateral flow begins at time t = 0. Clearly (3-25) and (3-26) both reduce to

$$c_t = \frac{qg}{2Bc} \tag{4-1}$$

because $w = 2c$ from (3-28), and $A = By = \frac{Bc^2}{g}$ in any cross section perpendicular to the side walls.

We use the finite difference expression (A-3) to give

$$c_a^{b+1} = c_a^b + \frac{qg\Delta t}{2Bc} \tag{4-2}$$

If we evaluate the variable c at $(a\Delta s,\ b\Delta t)$, we get, for the initial time step

$$c_a^1 = c_a^0 + \frac{qg\Delta t}{2Bc_a^0} \tag{4-3}$$

For the second time step

$$c_a^2 = c_a^1 + \frac{qg\Delta t}{2Bc_a^1} \tag{4-4}$$

Subtracting (4-3) from (4-4)

$$c_a^2 - c_a^1 = (c_a^1 - c_a^0)(1 - \frac{qg\Delta t}{2Bc_a^1 c_a^0}) \tag{4-5}$$

Similarly,

$$c_a^3 - c_a^2 = (c_a^2 - c_a^1)(1 - \frac{qg\Delta t}{2Bc_a^2 c_a^1}) = (c_a^1 - c_a^0)(1 - \frac{qg\Delta t}{2Bc_a^1 c_a^0})(1 - \frac{qg\Delta t}{2Bc_a^2 c_a^1})$$

$$c_a^{b+1} - c_a^b = (c_a^1 - c_a^0)(1 - \frac{qg\Delta t}{2Bc_a^1 c_a^0})(1 - \frac{qg\Delta t}{2Bc_a^2 c_a^1})\ldots\ldots(1 - \frac{qg\Delta t}{2Bc_a^b c_a^{b-1}}) \tag{4-6}$$

Note that $$1 - \frac{qg\Delta t}{2Bc_a^b c_a^{b-1}} = 1 - \frac{qg\Delta t}{2B(c_a^{b-1})^2 + qg\Delta t} \tag{4-7}$$

which is positive for all b, $\Delta t$ if q is positive.

For convenience we define mean values $c_m^b$ such that for all b

$$\left[1 - \frac{qg\Delta t}{2B(c_m^b)^2}\right]^b = \left[1 - \frac{qg\Delta t}{2Bc_a^1 c_a^0}\right]\left[1 - \frac{qg\Delta t}{2Bc_a^2 c_a^1}\right]\ldots\ldots\left[1 - \frac{qg\Delta t}{2Bc_a^b c_a^{b-1}}\right] \tag{4-8}$$

Thus, using (4-6)

$$c_a^{b+1} = c_a^1 + (c_a^1 - c_a^0)\left[\left(1 - \frac{qg\Delta t}{2B(c_m^1)^2}\right) + \left(1 - \frac{qg\Delta t}{2B(c_m^2)^2}\right)^2 + \ldots\ldots + \left(1 - \frac{qg\Delta t}{2B(c_m^b)^2}\right)^b\right] \quad (4\text{-}9)$$

Clearly the stability properties for the formulation depend on the factor $1 - \frac{qg\Delta t}{2B(c_m^b)^2}$, which can be called the amplification factor.

If we evaluate the variable c in (4-2) at $(a\Delta s, (b+1)\Delta t)$ we get in place of (4-5)

$$(c_a^2 - c_a^1)\left(1 + \frac{qg\Delta t}{2Bc_a^1 c_a^2}\right) = c_a^1 - c_a^0$$

Thus, with appropriate modification of the definition of $c_m^b$, we get by the same argument an amplification factor of

$$\left(1 + \frac{qg\Delta t}{2B(c_m^b)^2}\right)^{-1}$$

If q is positive, that is, a lateral inflow, the series in (4-9) will converge monotonically for all $\Delta t$. Now the calculated value $c_a^1 - c_a^0$ will almost certainly differ from $c(a\Delta s, \Delta t) - c(a\Delta s, 0)$ by some small initial error which will produce cumulative errors in a converging series in successive values of $c_a^{b+1}$. The true solution c(t) increases as t increases because the depth must increase with a lateral inflow, so the relative error must ultimately decrease and the solution is stable.

If q is negative, that is, a lateral outflow, the series in (4-9) will diverge monotonically for all $\Delta t$. As $(c_a^1 - c_a^0)$ is negative in this case, from (4-3), $c_m^b$ will decrease as t increases, so that the nonlinearity in $c_m^b$ will accentuate the divergence. Now from (4-1), $cc_t$ is constant, so we expect $c_t$ to increase as c decreases. However the initial error in $c_a^1 - c_a^0$ will produce cumulative errors in a diverging series in successive values of $c_a^{b+1}$. The relative error will increase even faster than the absolute error because the true c(t) will steadily decrease as t increases. Thus the solution for q negative is invalidated by monotonic instability which is characterised by uniform divergence between the true and calculated solutions with increasing t.

Note that these conclusions apply whether the variable c in (4-2) is evaluated at the backward or forward time step.

As a second simple example we take initially uniform flow in a river with a rectangular prismatic channel. $V_s$ and $w_s$ will be zero throughout in the first step, and hence (3-25) becomes

$$V_t + w_t = \frac{gz_s}{\eta} - \frac{gS_f}{\eta}$$

and (3-26) becomes

$$V_t - w_t = \frac{gz_s}{\eta} - \frac{gS_f}{\eta}$$

Thus we must have $w_t = 0$ in the first step. We shall assume that $V_t$ is assessed at the end boundaries in the same way as in the remainder of the channel. $V_s$ and $w_s$, and hence $w_t$, therefore retain their value of zero, so that the depth and hence the hydraulic radius R remain constant. Thus, using the Manning formula (2-40), our equations become

$$V_t = K_1 - K_2 V^2 \qquad (4\text{-}10)$$

where $K_1$ is a positive constant, and $K_2$ is a constant which takes the same sign as V.

We use the finite difference expression (A-3) to give

$$V_a^{b+1} = V_a^b + K_1 \Delta t - K_2 \Delta t V^2 \qquad (4\text{-}11)$$

Now it is a common stability test to check whether a difference solution can retain a steady state after an initial small disturbance has been introduced. For truly uniform flow, of course,

$$K_1 = K_2 V^2 \qquad (4\text{-}12)$$

We shall assume, however, that we evaluate V in (4-11) such that a small error is introduced, so that V does not exactly satisfy (4-12). The presence of rounding errors in all but exceptional numerical calculations makes this a reasonable assumption. We now investigate the circumstances in which that error grows until it dominates the solution.

If we evaluate V in (4-11) at $(a\Delta s, b\Delta t)$, we get, for the initial time step

$$V_a^1 = V_a^0 + K_1 \Delta t - K_2 \Delta t (V_a^0)^2 \qquad (4\text{-}13)$$

For the second time step

$$V_a^2 = V_a^1 + K_1 \Delta t - K_2 \Delta t (V_a^1)^2 \qquad (4\text{-}14)$$

Subtracting (4-13) from (4-14)

$$V_a^2 - V_a^1 = (V_a^1 - V_a^0)\left[1 - K_2\Delta t(V_a^1 + V_a^0)\right] \qquad (4\text{-}15)$$

Similarly, if we evaluate V in (4-11) at $(a\Delta s,(b+1)\Delta t)$ we get

$$(V_a^2 - V_a^1)\left[1 + K_2\Delta t(V_a^2 + V_a^1)\right] = V_a^1 - V_a^0 \qquad (4\text{-}16)$$

Clearly, by the same argument as before with appropriate definitions of $V_m^b$, we get an amplification factor of $1 - 2K_2\Delta tV_m^b$ when V is evaluated at the backward time step, and $(1 + 2K_2\Delta tV_m^b)^{-1}$ when V is evaluated at the forward time step. Now $K_2V_m^b$ is always positive because $K_2$ takes the same sign as V, so that $(1 + 2K_2\Delta tV_m^b)^{-1}$ is less than unity for all $\Delta t$. Thus the initial error $V_a^1 - V_a^0$ will be rapidly damped out if V in (4-11) is evaluated at the forward time step, and the solution will be stable.

Similarly the solution will be stable if V is evaluated at the backward time step provided

$$1 - 2K_2\Delta tV_m^b \geqslant -1 \qquad (4\text{-}17)$$

If the condition (4-17) is infringed, however, $V_a^{b+1} - V_a^b$ will uniformly increase in magnitude, alternating between positive and negative, so that the error quickly dominates the solution. Thus the solution in this case is invalidated by oscillatory instability which is characterised by rapidly increasing oscillations in the difference solution as t increases.

Note that this form of instability can be avoided either by limiting the time step $\Delta t$, or by using a different formulation, for instance, evaluating V at the forward time step.

### 4.3 Physical Instability

Uniform flow becomes physically unstable at high velocities and forms a series of roll waves. A solution can hardly be numerically stable if it is physically unstable whatever the formulation and however small the value of $\Delta t$ used. We would therefore expect a satisfactory stability theory to predict instability at the point established by existing physical stability theory when the Vedernikov number (Ven te Chow (1959) p.210) exceeds unity. Ven te Chow's definition of the Vedernikov number may be expressed as

$$V.N. = m(1 - R\frac{dP}{dA})F \tag{4-18}$$

where m is the exponent defined in (2-53) or (2-54) while R, P, A and F have their standard meanings for this thesis.

Clearly, in the case of a wide rectangular channel, this stability criterion becomes

$$mF \leqslant 1 \tag{4-19}$$

### 4.4 Stability Investigation with Finite Mesh Lengths - Introduction

The application of Fourier series stability analysis methods to nonlinear systems of difference equations is justifiable only if these equations can be linearised in some way. Quasi-linear equations, such as the Overland Flow equations, can be treated as "locally linear" in a small region of the solution, and this fact is used by Richtmyer in assessing the stability properties of solutions as the mesh lengths $\Delta t$ and $\Delta s$ tend to zero, as shown in Appendix A. However, the crucial feature of such linearisation is not the requirement that the mesh lengths be infinitesimal, but the requirement that the variations over a few mesh lengths in the dependent variables be an order of magnitude smaller than the corresponding variables themselves.

Thus we shall approach the Fourier series stability analysis of quasi-linear equations from a new viewpoint, assuming only that the difference solution varies gradually so that variations from the locally constant zero order solution are of the first order (small). In introducing this approach, we use as an example the staggered finite difference scheme (Lax, 1954), in the form recommended by Stoker (1957) p.481, because the stability analysis is relatively simple in this case. For simplicity we assume that the channel is wide, rectangular and prismatic, and that $\beta = \eta = 1$. We also take $U \equiv V$, which from (2-20) is the same as assuming $u = 0$ everywhere, and set $\cos \phi = 1$.

Hence we take $A = Bc^2/g$, B constant, $H_s = 0$, and $R = y$.

Thus, using (3-28) with $k = 1$, (3-25) and (3-26) become

$$V_t + 2c_t + (V + c)(V_s + 2c_s) = gz_s - gS_f - \frac{qVg}{Bc^2} + \frac{qg}{Bc} \tag{4-20}$$

$$V_t - 2c_t + (V - c)(V_s - 2c_s) = gz_s - gS_f - \frac{qVg}{Bc^2} - \frac{qg}{Bc} \tag{4-21}$$

We now form the "staggered scheme" finite difference expressions, following the notation of Appendix A.

$$c_t = \frac{2c_a^{b+1} - (c_{a-1}^b + c_{a+1}^b)}{2\Delta t} \qquad V_t = \frac{2V_a^{b+1} - (V_{a-1}^b + V_{a+1}^b)}{2\Delta t} \tag{4-22}$$

$$c_s = \frac{c_{a+1}^b - c_{a-1}^b}{2\Delta s} \qquad V_s = \frac{V_{a+1}^b - V_{a-1}^b}{2\Delta s} \tag{4-23}$$

Since we propose to use the same space differences (4-23) in (4-20) and (4-21) it is convenient to subtract, then add (4-20) and (4-21), giving

$$c_t + \frac{c}{2}V_s + Vc_s = \frac{qg}{2Bc} \tag{4-24}$$

$$V_t + VV_s + 2cc_s = gz_s - gS_f - \frac{qVg}{Bc^2} \tag{4-25}$$

Now if we wish to write the equations of first variation corresponding with (4-24) and (4-25), we must evaluate the non-homogeneous portion of the equations, that is, the R.H.S.

We say $$\frac{qg}{2Bc_a^b} = \frac{qg}{2B(c+\mathring{c}_a^b)} = \frac{qg}{2Bc}\left(1 - \frac{\mathring{c}_a^b}{c}\right)$$

because $\mathring{c}_a^b$ is a small first order variation of the zero order "local constant" c, and we are dropping all powers and multiples of first order quantities. In the staggered scheme we evaluate the quantities c and V in (4-24) and (4-25) as $\frac{1}{2}(c_{a-1}^b + c_{a+1}^b)$ and $\frac{1}{2}(V_{a-1}^b + V_{a+1}^b)$ so that

$$\frac{qg}{2B(c + \frac{1}{2}(\mathring{c}_{a-1}^b + \mathring{c}_{a+1}^b))} = \frac{qg}{2Bc}\left(1 - \frac{\mathring{c}_{a-1}^b + \mathring{c}_{a+1}^b}{2c}\right) \tag{4-26}$$

Evaluating $S_f$ by (2-53), we have

$$S_{fa}^{\ b} = \left[\frac{\frac{1}{2}(V_{a-1}^{b} + V_{a+1}^{b})g^{m}}{D\left(\dfrac{c_{a-1}^{b} + c_{a+1}^{b}}{2}\right)^{2m}}\right]^{1/j}$$

$$= S_f\left[1 + \frac{\mathring{V}_{a-1}^{b} + \mathring{V}_{a+1}^{b}}{2jV} - m\,\frac{(\mathring{c}_{a-1}^{b} + \mathring{c}_{a+1}^{b})}{jc}\right] \qquad (4\text{-}27)$$

by the same reasoning, where $S_f$ is now a function of the local constants c and V and is therefore also a local constant. Likewise

$$\frac{\frac{1}{2}q(V_{a-1}^{b} + V_{a+1}^{b})g}{B\left(\dfrac{c_{a-1}^{b} + c_{a+1}^{b}}{2}\right)^{2}} = \frac{qgV}{Bc^{2}}\left[1 + \frac{\mathring{V}_{a-1}^{b} + \mathring{V}_{a+1}^{b}}{2V} - \frac{\mathring{c}_{a-1}^{b} + \mathring{c}_{a+1}^{b}}{c}\right] \qquad (4\text{-}28)$$

Substituting (4-22), (4-23), (4-26), (4-27) and (4-28) into (4-24) and (4-25), multiplying by $\Delta t$, rewriting each difference term in terms of variation quantities, and dropping second order quantities, we get

$$\mathring{c}_{a}^{b+1} = \frac{\mathring{c}_{a-1}^{b} + \mathring{c}_{a+1}^{b}}{2} - \frac{\Delta t}{2\Delta s}\left[\frac{c}{2}\,(\mathring{V}_{a+1}^{b} - \mathring{V}_{a-1}^{b}) + V(\mathring{c}_{a+1}^{b} - \mathring{c}_{a-1}^{b})\right]$$
$$- \frac{qg\Delta t}{2Bc}\,\frac{\mathring{c}_{a-1}^{b} + \mathring{c}_{a+1}^{b}}{2c} + \frac{qg\Delta t}{2Bc} \qquad (4\text{-}29)$$

$$\mathring{V}_{a}^{b+1} = \frac{\mathring{V}_{a-1}^{b} + \mathring{V}_{a+1}^{b}}{2} - \frac{\Delta t}{2\Delta s}\left[V(\mathring{V}_{a+1}^{b} - \mathring{V}_{a-1}^{b}) + 2c(\mathring{c}_{a+1}^{b} - \mathring{c}_{a-1}^{b})\right]$$
$$- gS_f\Delta t\left[\frac{\mathring{V}_{a-1}^{b} + \mathring{V}_{a+1}^{b}}{2jV} - m\frac{\mathring{c}_{a-1}^{b} + \mathring{c}_{a+1}^{b}}{jc}\right] - \frac{qgV\Delta t}{Bc^{2}}\left[\frac{\mathring{V}_{a-1}^{b} + \mathring{V}_{a+1}^{b}}{2V}\right.$$
$$\left. - \frac{\mathring{c}_{a-1}^{b} + \mathring{c}_{a+1}^{b}}{c}\right] + gz_s\Delta t - gS_f\Delta t - \frac{qgV\Delta t}{Bc^{2}} \qquad (4\text{-}30)$$

Now the zero order terms $\frac{qg\Delta t}{2Bc}$ in (4-29) and $(z_s - S_f - \frac{qV}{Bc^2})g\Delta t$ in (4-30) are the products of locally constant quantities and $g\Delta t$, and are also locally constant. This means that they will contribute a steady increase or decrease to $\overset{\circ}{c}{}_a^{b+1}$ and $\overset{\circ}{V}{}_a^{b+1}$ over a few time steps, unaffected by any incipient instability in the solution. As we may assume that we commence with physically realistic initial conditions prior to the growth of any instability, it does not appear that the growth of errors associated with instability arises from the zero order terms.

If we use the Fourier series (A-12), the locally constant zero order terms will all be included in the constant term ($p = 0$) of the Fourier series and will not enter the harmonics at all. This reinforces the above conclusion that it does not appear that instability is associated with the zero order terms, as these cannot cause the unbounded amplification of a harmonic which is the symptom of instability in the Fourier series method. We may therefore drop all terms of zero order when we investigate the growth of the harmonics.

## 4.5 Fourier Series Representation

Using the Fourier series (A-12) and cancelling throughout by $e^{ipa\Delta s}$ as in Appendix A, (4-29) and (4-30) transform to

$$\overset{*}{c}{}^{b+1} = \left[\frac{e^{-ip\Delta s} + e^{ip\Delta s}}{2} - \frac{V\Delta t}{2\Delta s}(e^{ip\Delta s} - e^{-ip\Delta s}) - \frac{qg\Delta t}{4Bc^2}(e^{-ip\Delta s} + e^{ip\Delta s})\right]\overset{*}{c}{}^{b} - \frac{c\Delta t}{4\Delta s}(e^{ip\Delta s} - e^{-ip\Delta s})\overset{*}{V}{}^{b} \qquad (4\text{-}31)$$

$$\overset{*}{V}^{b+1} = \Big[ - \frac{c\Delta t}{\Delta s}(e^{ip\Delta s} - e^{-ip\Delta s}) + \frac{gS_f\Delta t}{jc} m(e^{-ip\Delta s} + e^{ip\Delta s})$$
$$+ \frac{qgV\Delta t}{Bc^3}(e^{-ip\Delta s} + e^{ip\Delta s}) \Big] \overset{*}{c}^{b} + \Big[ \frac{e^{-ip\Delta s} + e^{ip\Delta s}}{2}$$
$$- \frac{V\Delta t}{2\Delta s}(e^{ip\Delta s} - e^{-ip\Delta s}) - \frac{gS_f\Delta t}{2jV}(e^{-ip\Delta s} + e^{ip\Delta s})$$
$$- \frac{qg\Delta t}{2Bc^2}(e^{-ip\Delta s} + e^{ip\Delta s}) \Big] \overset{*}{V}^{b} \qquad (4\text{-}32)$$

for all $p \neq 0$. We define W and Z as in (A-18) and write $\theta$ for $p\Delta s$, allowing $\theta$ to vary between $-\pi$ and $\pi$ as in Appendix A. Note that $\theta$ can take the value zero even though p cannot, as $p\Delta s$ may be $\pm 2\pi$, $\pm 4\pi$ etc.

Let

$$\frac{gS_f}{2jV} = \sigma \qquad \frac{qg}{2Bc^2} = \gamma \qquad (4\text{-}33)$$

and with the Froude Number $F = V/c$, (4-31) and (4-32) become

$$\overset{*}{c}^{b+1} = \Big[ \cos\theta \; (1 - \gamma\Delta t) - iW \sin\theta \Big] \overset{*}{c}^{b} - \frac{iZ \sin\theta}{2} \overset{*}{V}^{b} \qquad (4\text{-}34)$$

$$\overset{*}{V}^{b+1} = \Big[ -2iZ \sin\theta + (4mF\sigma\Delta t + 4F\gamma\Delta t) \cos\theta \Big] \overset{*}{c}^{b} + \Big[ \cos\theta \; (1 - 2\gamma\Delta t - 2\sigma\Delta t) - iW \sin\theta \Big] \overset{*}{V}^{b} \qquad (4\text{-}35)$$

Note that W, Z, $\sigma\Delta t$, $\gamma\Delta t$ and F are all dimensionless locally constant quantities.

## 4.6 Eigenvalues of the Amplification Matrix

The amplification matrix $M_a$ of the system (4-34) and (4-35) is therefore

$$M_a = \begin{bmatrix} (1-\gamma\Delta t)\cos\theta - iW\sin\theta & \dfrac{-iZ\sin\theta}{2} \\ -2iZ\sin\theta + (m\sigma+\gamma)4F\Delta t\cos\theta & (1-2(\sigma+\gamma)\Delta t)\cos\theta - iW\sin\theta \end{bmatrix} \qquad (4\text{-}36)$$

and its eigenvalues are

$$\lambda = \cos\theta - iW\sin\theta - (\sigma + \frac{3\gamma}{2})\Delta t\cos\theta \pm \left[(\sigma+\frac{\gamma}{2})^2\Delta t^2\cos^2\theta - 2iZ\sin\theta\cos\theta F\Delta t(m\sigma+\gamma) - Z^2\sin^2\theta\right]^{\frac{1}{2}}$$

We set

$$\sigma + \frac{\gamma}{2} = L_1 \qquad F(m\sigma+\gamma) = L_2 \tag{4-37}$$

Therefore

$$\lambda = \cos\theta - iW\sin\theta - (L_1+\gamma)\Delta t\cos\theta \pm \left[L_1{}^2\Delta t^2\cos^2\theta - 2L_2\Delta t\cos\theta\, iZ\sin\theta - Z^2\sin^2\theta\right]^{\frac{1}{2}} \tag{4-38}$$

We seek to evaluate the square root as

$$\epsilon L_1\Delta t\cos\theta - i\zeta Z\sin\theta$$

Thus we set

$$(\epsilon L_1\Delta t\cos\theta - i\zeta Z\sin\theta)^2 = L_1{}^2\Delta t^2\cos^2\theta - 2L_2\Delta t\cos\theta iZ\sin\theta - Z^2\sin^2\theta \tag{4-39}$$

We equate real and imaginary parts of (4-39)

$$L_1{}^2\Delta t^2\cos^2\theta(1-\epsilon^2) - Z^2\sin^2\theta(1-\zeta^2) = 0 \tag{4-40}$$

$$2Z\Delta t\sin\theta\cos\theta(L_2 - \epsilon\zeta L_1) = 0 \tag{4-41}$$

and thus we can evaluate $\epsilon$ and $\zeta$.

## 4.7 Bounds on the Eigenvalues

From (4-38) and (4-39) we get

$$\lambda = \cos\theta\,(1 - (L_1 + \gamma \mp \epsilon L_1)\Delta t) - i\sin\theta\,(W \pm \zeta Z) \tag{4-42}$$

$$|\lambda|^2 = \cos^2\theta\,(1 - (L_1 + \gamma \mp \epsilon L_1)\Delta t)^2 + \sin^2\theta\,(W \pm \zeta Z)^2 \tag{4-43}$$

We are interested in the maxima of $|\lambda|^2$ which we can obtain by varying $\theta$. Hence we evaluate

$$\frac{d|\lambda|^2}{d\theta} = -2\cos\theta\sin\theta(1 - (L_1 + \gamma \mp \epsilon L_1)\Delta t)^2 + 2\sin\theta\cos\theta(W \pm \zeta Z)^2 \tag{4-44}$$

showing that there are turning points at $\sin\theta = 0$ and $\cos\theta = 0$. In fact the magnitude of any function in real constants and $e^{i\theta}$ will have turning points with respect to variation in $\theta$ when $\sin\theta = 0$, because the real part will comprise constant terms, terms in $\cos\theta$, and terms in $\cos^2\theta$, while the imaginary part will always be terms in $\sin^2\theta$. Differentiation w.r.t. $\theta$ must therefore give a factor of $\sin\theta$ in all the terms retained, which implies turning points when $\sin\theta = 0$. We use this in the next chapter.

Differentiating again,

$$\frac{d^2|\lambda|^2}{d\theta^2} = 2(\cos^2\theta - \sin^2\theta)\left[(W\pm\zeta Z)^2 - (1 - (L_1+\gamma\mp\epsilon L_1)\Delta t)^2\right] \qquad (4\text{-}45)$$

Now from (4-44) we get turning points in three cases:

1. $\cos\theta = 0$
2. $\sin\theta = 0$
3. $(W \pm \zeta Z)^2 = (1 - (L_1 + \gamma \mp \epsilon L_1)\Delta t)^2$

From (4-45) we see that if $(W \pm \zeta Z)^2 > (1 - (L_1 + \gamma \mp \epsilon L_1)\Delta t)^2$, the maxima occur in Case 1.

In this case, from (4-39), $\zeta^2 = 1$, so that from (4-43)

$$|\lambda|^2 \leqslant (W + Z)^2 \qquad (4\text{-}46)$$

If $(W \pm \zeta Z)^2 < (1 - (L_1 + \gamma \mp \epsilon L_1)\Delta t)^2$ the maxima occur in Case 2. In this case $\epsilon^2 = 1$, so that from (4-43)

$$|\lambda|^2 \leqslant (1 - (L_1 + \gamma \mp L_1)\Delta t)^2 \qquad (4\text{-}47)$$

In Case 3, we see from (4-43) that

$$|\lambda|^2 = (W \pm \zeta Z)^2 (\cos^2\theta + \sin^2\theta) \qquad (4\text{-}48)$$

which is independent of $\theta$. In this case we must investigate the behaviour of $\zeta$ from (4-40) and (4-41). $\zeta$ is defined by these equations except when $\sin\theta\cos\theta = 0$.

When $\sin\theta = 0$, $\epsilon^2 = 1$ from (4-39) and hence

$$(W \pm \zeta Z)^2 = (1 - (L_1 + \gamma \mp L_1)\Delta t)^2$$

so that Case 3 leads to the equality in (4-47) for $\sin\theta = 0$. Similarly, when $\cos\theta = 0$, Case 3 gives the equality in (4-46). If $\sin\theta\cos\theta \neq 0$, (4-41) gives $\epsilon = L_2/\zeta L_1$, and also $\zeta \neq 0$. Thus, substituting in (4-40)

$$L_1^2\Delta t^2\cos^2\theta - \frac{L_2^2}{\zeta^2}\Delta t^2\cos^2\theta - Z^2\sin^2\theta\,(1 - \zeta^2) = 0$$

$$Z^2\sin^2\theta\,\zeta^4 + (L_1^2\Delta t^2\cos^2\theta - Z^2\sin^2\theta)\zeta^2 - L_2^2\Delta t^2\cos^2\theta = 0 \qquad (4\text{-}49)$$

This has roots $\zeta_1^2$, $\zeta_2^2$, where

$$\zeta_1^2\zeta_2^2 = \frac{-L_2^2\Delta t^2\cos^2\theta}{Z^2\sin^2\theta}$$

That is, one of $\zeta_1$, $\zeta_2$ is real and the other imaginary. Taking the real root $\zeta_r$ and rearranging (4-49),

$$(L_1^2 - L_2^2)\Delta t^2\cos^2\theta = (L_1^2\Delta t^2\cos^2\theta + \zeta_r^2 Z^2\sin^2\theta)(1 - \zeta_r^2) \qquad (4\text{-}50)$$

Thus if $L_1^2 \geqslant L_2^2$, we have $\zeta_r^2 \leqslant 1$.

That is, if $(\sigma + \frac{\gamma}{2})^2 \geqslant F^2(m\sigma + \gamma)^2$, $\quad \zeta_r^2 \leqslant 1$

using (4-37). Neglecting $\gamma$, see below, this condition becomes

$$m^2F^2 \leqslant 1 \qquad (4\text{-}51)$$

If (4-51) is compared with (4-19), it is plain that $\zeta_r^2$ will exceed unity if and only if the flow is physically unstable.

We are entitled to neglect $\gamma$ above because (4-19) was derived for a flow without lateral inflow, and comparison between

(4-51) and (4-19) is possible only under comparable conditions. The coincidence between (4-51) and (4-19) is no accident, as we will show in the next chapter under quite general conditions that a physically unstable solution is also numerically unstable.

Meanwhile, from (4-48) and (A-18), in Case 3 we have

$$|\lambda|^2 = (V \pm \zeta c)^2 \left(\frac{\Delta t}{\Delta s}\right)^2 \tag{4-52}$$

We compare (4-52) with the Courant condition which can be written

$$(V \pm c)^2 \left(\frac{\Delta t}{\Delta s}\right)^2 \leqslant 1 \tag{4-53}$$

If the flow is physically unstable, $\zeta^2$ exceeds unity in (4-52) and the von Neumann stability criterion (A-14) may be violated even if the Courant condition is satisfied, while for physically stable flows satisfaction of the Courant condition will guarantee satisfaction of the von Neumann stability criterion by $\lambda$ as evaluated for Case 3 from (4-52). The form of (4-52) is therefore reminiscent of the remark by Henderson (1966) p.369 that the "kinematic wave" velocity (roughly the velocity at which the main body of a flood wave moves) will be greater or less than the dynamic wave velocity V + c according as the flow is physically unstable or stable respectively.

Assuming the flow is physically stable, inequality (4-46) covers equation (4-52), so that in fact inequality (4-46) covers all Case 3 for $\sin\theta \neq 0$.

Hence we have established limits on $|\lambda|^2$ for all possibilities and can say, from (4-46) and (4-47)

$$|\lambda|^2 \leq \max\left[(W + Z)^2,\ (1 - \gamma\Delta t)^2,\ (1 - 2(\sigma + \gamma)\Delta t)^2\right] \qquad (4\text{-}54)$$

where we have used (4-37) to evaluate $2L_1 + \gamma$.

## 4.8 Stability Criteria

Having established bounds on the eigenvalues, we must decide on a stability criterion. Now our analysis should not differ from Richtmyer's methods in any way as $\Delta t$ tends to zero, because all terms $O(\Delta t)$, i.e. the terms* which arise from the R.H.S. of (4-24) and (4-25) after their difference representations are multiplied by $\Delta t$. Hence we can use the von Neumann necessary condition for stability (A-14) as $\Delta t$ tends to zero. In this case we get, from the first limit of (4-54), the Courant condition (4-53) coinciding with the von Neumann necessary condition for stability (provided, as we have seen, that the solution describes a physically stable flow). It is not clear whether the von Neumann necessary condition is also sufficient for stability as $\Delta t$ tends to zero, but if we consider

$$\Delta t = \frac{2}{\sigma + \gamma} \qquad (4\text{-}55)$$

We have, from the third limit of (4-54)

$$|\lambda|^2 \leq (-3)^2$$

Now a solution in which it is possible for some harmonic not only to reverse in sign at each time step but treble in magnitude at the same time cannot be regarded as stable. On the

---

* Because the R.H.S. of (4-24) and (4-25) makes these equations non-homogeneous, we shall hereafter describe these terms as "non-homogeneous terms" whether we are referring to such terms in the partial differential equations or to their representation in a finite difference scheme.

other hand, if we impose the restriction

$$|\lambda| \leqslant 1 \tag{4-56}$$

for all $\Delta t$, we may reasonably expect any disturbances in the solution not to increase, so that we take (4-56) as our sufficient condition for stability.

This condition is the one generally used throughout the literature (e.g. Smith (1965) p.71) as it is intuitively necessary and sufficient. However, as pointed out by Richtmyer (1957), Chapter IV, in a case where the solution itself grows exponentially, we would expect a legitimate exponential growth of the stability errors, in which case the errors might remain tolerably small by comparison with the solution. Hence the solution would for practical purposes be stable. Another possibility is that $|\lambda|$ might just exceed unity so that any amplification of errors would be so gradual that the final error at the end of the numerical solution might still be small enough to be tolerable. Hence the von Neumann condition is the proper necessary condition for stability as it is more generous than (4-56) by just enough to permit "acceptable" growth in the stability errors.

Thus for finite mesh lengths we can suppose that the numerical solution is stable if it satisfies condition (4-56) and unstable if it violates the von Neumann condition. If it satisfies the von Neumann condition but not condition (4-56), we must decide whether the error may become intolerably large over the time period in which (4-56) is violated. In general the error will indeed grow rapidly to intolerable proportions because of the large number of time steps usually handled by computers in automatic solutions, so that in this thesis we shall regard any violations of (4-56) as indicating instability. Note that we have in effect substituted for the definition of stability in the previous chapter the more practical requirement that the stability error should remain an order of magnitude less than the solution.

We have assumed in (4-55) that a value of $\Delta t$ as large as $\frac{2}{\sigma + \gamma}$ might be used in an actual solution. We prove this by an example, with $q = 0$, $S_f = .001$, $V = 4$ ft/sec, $j = \frac{1}{2}$, $g = 32$ ft/sec$^2$. Using (4-33)

$$\Delta t = \frac{2 \times 4}{.032} \text{ seconds} = 250 \text{ seconds, or approx 4 minutes.}$$

Now this value of $\Delta t$ is a common magnitude for flood routing in rivers, for instance by Stoker (1957) Chapter 11, and hence stability criteria involving the magnitude of $\Delta t$ and condition (4-56) are of great importance. Thus we suggest that all of the failures of explicit finite difference schemes noted in Section 4.1 might well have been caused by the failure to recognise a limit on the magnitude of $\Delta t$ in addition to the Courant stability condition.

We now derive a set of sufficient conditions for stability from (4-54) and (4-56), assuming

$$F^2(m\sigma + \gamma)^2 \leq (\sigma + \frac{\gamma}{2})^2 \qquad (4\text{-}57)$$

which is apparently a criterion for physical stability. The conditions, which must all independently be satisfied, are

$$(V + c)\frac{\Delta t}{\Delta s} \leqslant 1 \tag{4-58}$$

$$\gamma \geqslant 0 \tag{4-59}$$

$$(\sigma + \gamma)\Delta t \leqslant 1 \tag{4-60}$$

These conditions of course apply only to the staggered scheme for which they were derived, but serve as an example of the results of this method of Fourier series analysis with finite mesh lengths. Note that (4-59) predicts the instability for all $\Delta t$ indicated for q negative by the first example of Section 4.2, while we may show that (4-17) and (4-60) coincide as follows. (4-17) can be rewritten, identifying $V_m^b$ with the locally constant V,

$$K_2 \Delta t V \leqslant 1$$

and

$$K_2 V = \frac{gS_f}{V} = \sigma$$

using (4-10), and (4-33) with $j = \frac{1}{2}$ for turbulent flow. Thus (4-17) and (4-60) coincide as $\gamma$ is taken as zero in (4-10).

## 4.9 Practical Use of Stability Criteria

Stoker drew attention (Section 4.1) to the problem of assessing the Courant condition prior to a solution, as the condition is expressed in terms of the unknown solution. Clearly, from (4-58), the Courant condition is still of great importance in the light of our extended stability analysis, and we now examine the particle velocity V + c, which is the factor in the Courant condition which must be predicted before a solution is carried out.

From (3-14), setting $A = Q/V$

$$c = \left[\frac{g \cos\phi \; Q}{\eta BV}\right]^{\frac{1}{2}}$$

Therefore
$$c^{\frac{3}{2}} = \left[\frac{g \cos\phi \; Q}{\eta BF}\right]^{\frac{1}{2}}$$

and hence

$$c = \left[\frac{g \cos\phi \; Q}{\eta BF}\right]^{\frac{1}{3}}$$

where F is again the Froude Number. Hence

$$V + c = \frac{(F + 1)}{F^{\frac{1}{3}}} \left(\frac{g \cos\phi}{\eta}\right)^{\frac{1}{3}} \left(\frac{Q}{B}\right)^{\frac{1}{3}} \qquad (4\text{-}61)$$

Now if a preliminary estimate shows $F < 0.1$, we can neglect V, and calculate c from (3-14). A good estimate of c should be possible, because in most cases we can assume

$$c \propto y^{\frac{1}{2}}$$

by consideration of simple channels, and hence an error in y will be at least halved in c.

The factor $(F+1)/F^{\frac{1}{3}}$ is almost constant for all F between 0.1 and 2.0, lying between 1.89 and 2.38, so if a preliminary estimate shows $F > 0.1$, as will occur in most cases, we can assume from (4-61) that

$$V + c \propto \left(\frac{Q}{B}\right)^{\frac{1}{3}} \qquad (4\text{-}62)$$

Now using (2-54)

$$\frac{Q}{B} = DR^m S_f{}^j \, \frac{A}{B} = \frac{DS_f{}^j}{L^m k} \, y^{m+1}$$

using (2-35) and (2-36).

$$Q = DS_f{}^j \frac{y^m}{L^m} Ey^k$$

using (2-37). Thus if we assume the channel is simple so that k, L and E are constant at any cross section, and also that $S_f$ remains constant, we have

$$y \propto Q^{\frac{1}{m+k}}$$

$$\frac{Q}{B} \propto Q^{\frac{m+1}{m+k}}$$

$$V + c \propto Q^{\frac{m+1}{3(m+k)}} \qquad (4\text{-}63)$$

As many channels approximate a parabola or trapezium in cross section, k is often about 3/2, so that

$$\frac{m+1}{3(m+k)} \approx \frac{1}{4} \qquad (4\text{-}64)$$

for turbulent flow, where $m = \frac{2}{3}$ in Manning's formula (2-40), and $m = \frac{1}{2}$ in the Chézy formula (2-39).

Thus a 100% overestimate in Q normally leads to less than a 20% overestimate in the particle velocity, so a generous estimate of the maximum discharge to be encountered in the solution is generally sufficinet information for an adequate safe estimate of the particle velocity V + c. This enables the choice of a suitable mesh ratio for a solution to be made automatically by a computer at the commencement of a solution, as the only input information needed, apart from geometrical information about the cross sections, is an estimate of the maximum discharge, and the velocity at this discharge. From these, A and hence B and c may be derived, and therefore V + c. It is not necessary to use

(4-61) to evaluate V + c therefore; the value of (4-61) lies in demonstrating that large errors are tolerable in the estimates of maximum discharge and velocity.

It is now also clear that even if the Courant condition is predominant, that is, if the desired $\Delta s$ is sufficiently small for any criteria on $\Delta t$ alone to be always satisfied when the Courant condition is satisfied, there is little point in altering the time interval $\Delta t$ at each step in the solution according to the value of the particle velocity as suggested by Fenzl (1965), because the changes in particle velocity V + c will generally be so small that savings in the number of time steps in a solution will hardly compensate for the extra computation involved in continually resetting the time step. In addition the use of constant time steps involves less programming effort and allows output to be produced at regular intervals throughout the solution.

The stability criterion (4-59) implies that satisfactory solutions cannot be obtained by the staggered scheme, of overland flow equations involving lateral outflows, and we show in the next chapter that this applies to all difference schemes based on a regular grid.

Criterion (4-60) requires an estimate of $\sigma$ and $\gamma$. In practice $\gamma$ is generally much smaller than $\sigma$, as will be shown in detail in Section 6.2. Meanwhile we shall check the condition (4-60) against the experience of Stoker (1957)

Pp.488-494 with the use of the staggered scheme in the case $\gamma = 0$.

From (4-33) and (4-60), we have

$$\frac{gS_f\Delta t}{2jV} \leqslant 1$$

Stoker gives a datum slope of about .0001 so that, assuming $S_f = z_s$, we get for $j = \frac{1}{2}$

$$\Delta t \leqslant \frac{10000}{g} V$$

That is, $\Delta t \leqslant 0.1\ V$ hours

approximately, where V is in miles per hour as used by Stoker.

Stoker unconsciously satisfied this criterion throughout, so he had no difficulties with the staggered finite difference scheme in contrast to his problems with the rectangular finite difference scheme. As we show in Chapter 6, these problems are also predictable by our extended stability analysis.

Now in most cases it should be possible to assume $S_f = z_s$ so that advance prediction of $S_f$ is possible; however there appears to be no way to predict the factor V in $\sigma$ with any accuracy and thus a finite difference scheme which has stability properties independent of $\sigma$ would be a great advantage. We now proceed to investigate the derivation of such a scheme.

# CHAPTER 5

## LIMITED STABILITY TESTS ON A GENERAL FINITE DIFFERENCE SCHEME

### 5.1 Introduction

In Chapter 4 we established a method of applying Fourier series stability analysis which gives more general results than stability analyses in use at present. This is the first time, to the knowledge of the writer,* that a Fourier series stability analysis has been performed on the finite difference representation of any hyperbolic non-homogeneous system of quasi-linear partial differential equations without first eliminating the non-homogeneous terms (See Section 5.8) by allowing the mesh lengths to tend to zero. Indeed it is only during the completion of the project leading to this thesis that the importance of non-homogeneous terms in stability analysis appears to have been recognised at all in the literature in English, when Vreugdenhil (1968) used Fourier series stability analysis to show that a non-homogeneous term, in a simple linear partial differential equation in one dependent variable, could introduce a stability restriction on the time increment in a difference solution.

While we could apply our method to any difference scheme to derive stability criteria for comparison, it is better to analyse a generalised finite difference scheme so that basic

---

* A similar stability analysis of a difference scheme by Liggett and Woolhiser (1968) has since been noted by the writer. However this analysis is incomplete in that the term $O(\Delta t^2)$ is dropped from the square root part of the equation corresponding with (4-38).

principles are not obscured as they may be in a trial and error series of analyses. In this way we should be able to establish fundamental stability requirements of difference schemes and select a scheme which best meets those requirements. Another advantage of generality is that the algebraic manipulation in these stability analyses is usually tedious and often intricate. Hence it is desirable to perform such manipulation in a general way as far as possible so that individual criteria may then be found in a straightforward manner.

In this chapter we apply Fourier series analysis to a general rectangular difference scheme applied to general overland flow equations and establish a number of stability criteria which are necessary to satisfy (4-56), in particular the requirement that for numerical stability the flow must be physically stable. In the following chapter we establish sufficient stability criteria, within the limits of our linearising assumptions, for a selected class of rectangular finite difference schemes. This results in the recommendation of a simple finite difference scheme with tractable stability properties.

## 5.2 Basic Assumptions

The method of Chapter 4 has three basic requirements. First, we must be able to write equations of first variation of our finite difference equations in such a way that they are linear in the first order variation terms. This is possible only if $\mathring{V}/V \ll 1$ and $\mathring{y}/y \ll 1$ at any point in the solution at which we wish to investigate stability. This is intuitively necessary over the major part of any acceptable finite difference solution, as the very replacement of differentials with differences cannot be regarded as valid if the solution exhibits variations, between adjacent grid points, of the same order as the solution itself. Exceptions to this are tolerated in special areas of a solution, such as the initial time step in which a dry channel is subjected to lateral inflow, and in these areas our analysis does not hold. However, all initial and boundary conditions should be examined separately, and numerical problems in these areas bear little relation to stability as defined in Chapter 3.

The second requirement is the restriction of our analysis to one cycle, two level finite difference schemes based on a rectangular mesh. By "one cycle" we mean a scheme which replaces the differential equations with a single finite difference formulation which is then solved repeatedly for each successive time step. By "two level" we mean that all functions of the dependent variables are expressed in the scheme in terms of the instantaneous solutions at only two times. Such a scheme may be explicit, as in Stoker (1957) Section 11.5, or implicit, as in Abbott and Ionescu (1967) p.99.

Multi-level schemes could possibly be handled by an extension of methods suggested by Richtmyer (1957) Chapter V, but such schemes are more cumbersome than one cycle two level schemes as they require special starting procedures, and extra computer storage to hold the solutions at the extra time levels.

Two cycle or iterative schemes require more intricate programming than one cycle two level schemes, so if the best one cycle two level rectangular finite difference schemes are found to be stable and consistent there is no reason to use more complicated rectangular schemes.

We postpone to Chapter 7 consideration of characteristic net finite difference schemes, to which this stability analysis does not apply.

The third requirement in generalising the methods of Chapter 4 arises from our need to deal with variations in only two dependent variables, so that we are required to relate $\mathring{w}$, $\mathring{R}$, $\mathring{A}$, $\mathring{y}$ and $\mathring{c}$ at any cross section by explicit locally constant relationships. We shall show this requirement may be satisfied if we assume that our general channel behaves like a simple channel when the depth varies only slightly from a locally constant value. We shall describe channels conforming to this assumption as "quasi-simple".

5.3 Quasi-Simple Channels

We define a quasi-simple channel as one in which k and L at a cross section, as defined by (2-35) and (2-36), undergo second order variations when a first order variation in depth is introduced at that section. This may be written, for any section s = const

$$\frac{\mathring{k}(y)}{k} \ll \frac{\mathring{y}}{y} \tag{5-1}$$

and

$$\frac{\mathring{L}(y)}{L} \ll \frac{\mathring{y}}{y} \tag{5-2}$$

We have written the variations $\mathring{k}(y)$, $\mathring{L}(y)$ to indicate that they are variations with y at a section, with s held constant. We are treating our geometrical parameters w, R, A, c, and hence B', k and L, as functions of y and s in the same way as introduced in Section 3.1. Thus, for instance, the total differential of A(y,s) is

$$dA = A_y dy + A_s ds$$

where $A_y dy$ applies for s constant and $A_s ds$ for y constant. Introducing variations, we write

$$\mathring{A} = A_y \mathring{y} + A_s \mathring{s}$$

or alternatively

$$\mathring{A} = \mathring{A}(y) + \mathring{A}(s)$$

Thus $\mathring{k}(y)$ tends to $k_y dy$ as $\mathring{y}$ tends to zero, and similarly $\mathring{L}(y)$ tends to $L_y dy$.

A trapezoidal channel is the most common practical example of a quasi-simple channel. For a trapezoidal cross section

$$A = E_1 y + E_2 y^2$$

where $E_1(s)$ is the bottom width and $E_2(s)$ is the mean side slope of the channel sides. From (2-35), using $B = A_y$,

$$k(y,s) = \frac{E_1 + 2E_2 y}{E_1 + E_2 y}$$

and hence

$$\frac{\mathring{k}(y)}{\mathring{y}} = k_y = \frac{E_1 E_2}{(E_1 + E_2 y)^2}$$

Thus $$\frac{\mathring{k}(y)}{k} = \frac{E_1E_2y}{(E_1+2E_2y)(E_1+E_2y)}\,\frac{\mathring{y}}{y}$$

Now $$\frac{\partial}{\partial y}\left[\frac{E_1E_2y}{(E_1+2E_2y)(E_1+E_2y)}\right] = \frac{E_1E_2(E_1{}^2-2E_2{}^2y^2)}{(E_1+2E_2y)^2(E+E_2y)^2}$$

and therefore when $y = \dfrac{E_1}{\sqrt{2}E_2}$, $\dfrac{\mathring{k}(y)}{k}\Big/\dfrac{\mathring{y}}{y}$ has its maximum value,

which is $$\frac{1}{1+2\sqrt{2}+2} = \frac{1}{5.83}$$

Therefore, for a trapezoidal channel

$$\frac{\mathring{k}(y)}{k} < \frac{1}{5.83}\,\frac{\mathring{y}}{y} \qquad (5\text{-}3)$$

which may be taken to satisfy (5-1).

Equation (5-2) is satisfied if equation (5-1) is satisfied and the ratio of P/B is approximately constant, as may be seen from (2-38). Thus using (5-3), we may say that a trapezoidal channel is quasi-simple provided that the flow is wide and shallow, or else sufficiently deep to approximate a triangle in cross-section. This excludes trapezoidal shapes which approach a deep narrow rectangle, but such shapes are rarely of practical interest and can be disregarded.

Because (5-1) and (5-2) are reasonably accurate in all likely trapezoidal shapes, it is reasonable to suppose that both also apply to most channel sections encountered in practice, so that our assumption that the general channel is quasi-simple should not greatly restrict the application of our analysis.

We now examine the relationships between $\mathring{c}(y)$, $\mathring{w}(y)$, $\mathring{A}(y)$ and $\mathring{R}(y)$ with $\mathring{y}$ at a section s = const.

From (3-14) and (2-35)

$$c = +\sqrt{\frac{g\cos\phi}{\eta}}\,\frac{y^{\frac{1}{2}}}{k^{\frac{1}{2}}}$$

$$c_y = +\sqrt{\frac{g\cos\phi}{\eta}}\left(\frac{1}{2k^{\frac{1}{2}}y^{\frac{1}{2}}} - \frac{y^{\frac{1}{2}}}{2k^{\frac{3}{2}}}k_y\right) = \frac{c}{2}\left(\frac{1}{y} - \frac{k_y}{k}\right)$$

Thus $$\frac{2\mathring{c}(y)}{c} = \frac{\mathring{y}}{y} - \frac{\mathring{k}(y)}{k}$$

$$= \frac{\mathring{y}}{y} \qquad (5\text{-}4)$$

for a quasi-simple channel.

From (3-22) we have, as w is a function of y alone,

$$\mathring{w}(y) = \sqrt{\frac{g\cos\phi\ k}{\eta\, y}}\ \mathring{y}$$

$$= ck\frac{\mathring{y}}{y}$$

From (5-4) $\mathring{w}(y) = 2k\mathring{c}(y)$ (5-5)

The relation at any section between $\mathring{A}(y)$ and $\mathring{y}$ is clearly, from (3-1)

$$\mathring{A}(y) = B\mathring{y} \qquad (5\text{-}6)$$

and between $\mathring{R}(y)$ and $\mathring{y}$, from (2-36)

$$\frac{\mathring{R}(y)}{R} = \frac{\mathring{y}}{y} - \frac{\mathring{L}(y)}{L}$$

$$= \frac{\mathring{y}}{y} \qquad (5\text{-}7)$$

for a quasi-simple channel, according to (5-2).

Note that these relationships (5-4), (5-5), (5-6) and (5-7) also apply to variations in depth with s if the quasi-simple channel is prismatic, because in this case k, L and B are functions of depth alone, and the depth can be taken as constant when such variations are small.

## 5.4 Physical Stability Criteria in Quasi-Simple Channels

We may simplify the factor $1 - R\frac{dP}{dA}$ in the definition of the Vedernikov number (4-18), which applies to prismatic channels, if we also assume the channel is quasi-simple. Thus

$$1 - R\frac{dP}{dA} = 1 - R\frac{d(A/R)}{dA}$$

$$= -RA\frac{d(\frac{1}{R})}{dR}\,\frac{dR}{dy}\,\frac{dy}{dA}$$

$$= \frac{A}{RLB}$$

$$= \frac{A}{By} \qquad (5\text{-}8)$$

using (2-36), (5-2) and dA/dy = B for a prismatic channel.

Thus the Vedernikov number stability criterion becomes simply

$$\left|mF\right| \leqslant \frac{By}{A} \qquad (5\text{-}9)$$

which for quasi-simple channels is more conveniently expressed, using (2-35)

$$\left|mF\right| \leqslant k \qquad (5\text{-}10)$$

The physical stability of flow in open channels was also investigated by Escoffier and Boyd (1962), who derived two criteria for stability

$$\frac{dV_o}{dw} \leqslant 1 \qquad (5\text{-}11)$$

and
$$\frac{dV_o}{dw} \geqslant -1 \qquad (5\text{-}12)$$

where w is as defined in (3-23) (although $\cos\phi$ and $\eta$ were taken to be unity), and $V_o$ is the normal velocity in the channel section defined by the Manning equation

$$V_o = C(z_s)^{\frac{1}{2}} R^{\frac{2}{3}}$$

Thus (see (2-40)), $V_o$ is the velocity at which the datum slope $z_s = S_f$.

We seek to prove that the Vedernikov number $\frac{mF}{k}$ is identical to $\frac{dV_o}{dw}$ in prismatic quasi-simple channels.

$$\frac{mF}{k} = \frac{mVA}{cBy} = \frac{mV}{y}\frac{A}{B}\sqrt{\frac{B\eta}{g\cos\phi\; A}} = \frac{mV}{y}\sqrt{\frac{A\eta}{g\cos\phi\; B}}$$

$$= \frac{mV}{y}\frac{dy}{dw} \qquad (5\text{-}13)$$

using (3-22).

$$\frac{dV_o}{dy} = C(z_s)^{\frac{1}{2}}\, \frac{2}{3}R^{-\frac{1}{3}}\, \frac{dR}{dy}$$

Therefore
$$\frac{dV_o}{dw} = \frac{2}{3}\frac{V_o}{R}\frac{dR}{dw} \qquad (5\text{-}14)$$

Now for Manning's Formula $m = \frac{2}{3}$ from (2-40) and (2-54), so that dividing (5-13) by (5-14) gives

$$\frac{mF}{k}\Big/\frac{dV_o}{dw} = \frac{V}{V_o}\frac{dy}{dR}\frac{R}{y} = 1 \qquad (5\text{-}15)$$

as expected, using (5-2) and the fact that Escoffier and Boyd assumed that $V = V_o$ throughout. We must evidently include the criterion, from (5-12)

$$-mF \leqslant k \qquad (5\text{-}16)$$

in (5-10) as indicated, because although m and k are always positive, F takes the same sign as the velocity (positive in the direction of s increasing), so that (5-16) is not automatically satisfied.

Thus for laminar flows, where m = 2, (5-10) becomes

$$|F| \leqslant k/2 \tag{5-17}$$

as was quoted in Chapter 3.

The criterion (5-9) therefore holds for all flows in quasi-simple prismatic channels, which may be "wide" or approximately triangular in cross section, as discussed in Section 5.3. Thus there should rarely be any necessity to discard the simple form of (5-9) in favour of the more complicated forms (4-18), derived by Vedernikov, or (5-11) and (5-12), derived by Escoffier and Boyd.

## 5.5 Variations of the Non-Homogeneous Terms

We are assuming that our geometrical parameters w, R, A, y and c can all be treated as locally constant in a small area of the solution, but these local constants bear constant relationships with each other only in prismatic channels. In non-prismatic channels a constant value of y, for instance, does not correspond with a constant value of A, as may be seen from equation (3-2). However, as shown in (3-5), the variation of A with s for y locally constant is a function of the channel geometry and the locally constant y. This can also be expressed by the total differential

$$dA(y,s) = A_y dy + A_s ds$$

or in the notation introduced in Section 5.3

$$\mathring{A} = \mathring{A}(y) + \mathring{A}(s)$$

where $\mathring{A}(s)$ is a function of fixed channel geometry and the locally constant y, and is therefore a locally constant quantity of the first order. We can write similar equations relating $\mathring{w}$ to $\mathring{w}(y)$, $\mathring{R}$ to $\mathring{R}(y)$, and $\mathring{c}$ to $\mathring{c}(y)$.

We shall now select $\mathring{c}$ as our principal depth variation because $\mathring{c}$ is normally used in simple stability analyses such as these in Appendix A and Chapter 4. We wish to express all variations of the non-homogeneous terms as terms in $\mathring{V}$, $\mathring{c}$ and $\mathring{f}(s)$, where we use $\mathring{f}(s)$ to denote <u>any</u> locally constant variation quantity which is independent of $\mathring{V}$ and $\mathring{y}$. Note that

$$\mathring{c}(y) = \mathring{c} - \mathring{c}(s) = \mathring{c} + \mathring{f}(s) \tag{5-18}$$

From (3-24) the non-homogeneous terms are $\frac{gz_s}{\eta}$, $\frac{gS_f}{\eta}$, $\frac{qU}{A\eta}$, $\frac{qc}{A}$ and $\frac{VH_s c}{A}$.

We take g and $\eta$ constant throughout, and q as specified independently, while $z_s$ and $H_s$ are regarded as at least locally constant.

Each variable will be evaluated at some central point, or else evaluated over several points and averaged, for each application of any difference scheme. We therefore study the variations introduced if we evaluate the variables at any point $(a\Delta s, b\Delta t)$. We shall write any variable, for instance area A, which is evaluated at $(a\Delta s, b\Delta t)$ as

$$A_a^b = A + \mathring{A}_a^b = A + \mathring{A}_a^b(y) + \mathring{f}(s)$$

using the unannotated variable for the locally constant value of the variable.

Thus, from (2-53)

$$S_{fa}^{\ b} = \left[\frac{V + \mathring{V}_a^b}{D(R+\mathring{R}_a^b)^m}\right]^{1/j}$$

$$= S_f\left[1 + \frac{\mathring{V}_a^b}{jV} - \frac{m\mathring{R}_a^b}{jR}\right]$$

$$= S_f\left[1 + \frac{\mathring{V}_a^b}{jV} - \frac{m(\mathring{R}_a^b(y) + \mathring{R}_a^b(s))}{jR}\right]$$

$$= S_f\left[1 + \frac{\mathring{V}_a^b}{jV} - \frac{2m(\mathring{c}_a^b - \mathring{c}_a^b(s))}{jc} - \frac{m\mathring{R}_a^b(s)}{jR}\right]$$

$$= S_f\left[1 + \frac{\mathring{V}_a^b}{jV} - \frac{2m\mathring{c}_a^b}{jc}\right] + \mathring{f}(s) \qquad (5\text{-}19)$$

using (5-7), (5-4) and (5-18).

We next examine U/A, assuming that U = V-u as in (2-20) and that u is specified independently in the same way as q.

Thus

$$\frac{U_a^b}{A_a^b} = \frac{V + \mathring{V}_a^b - u}{A + \mathring{A}_a^b}$$

$$= \frac{U}{A}\left(1 - \frac{2By\mathring{c}_a^b}{Ac}\right) + \frac{\mathring{V}_a^b}{A} + \mathring{f}(s) \qquad (5\text{-}20)$$

using (5-6) and (5-4).

Similarly

$$\frac{c_a^b}{A_a^b} = \frac{c}{A}\left[1 + \frac{\mathring{c}_a^b}{c}\left(1 - \frac{2By}{A}\right)\right] + \mathring{f}(s) \qquad (5\text{-}21)$$

and
$$\frac{V_a^b \, c_a^b}{A_a^b} = \frac{Vc}{A}\left[1 + \frac{\overset{\circ}{c}{}_a^b}{c}\left(1 - \frac{2By}{A}\right) + \frac{\overset{\circ}{V}{}_a^b}{V}\right] + \overset{\circ}{f}(s) \tag{5-22}$$

## 5.6 Fourier Series Representation

Equation (5-5) allows us to write

$$\overset{\circ}{w}{}_a^b = 2\frac{By}{A}\,\overset{\circ}{c}{}_a^b + \overset{\circ}{f}(s) \tag{5-23}$$

Thus by using equations (5-19), (5-20), (5-21), (5-22) and (5-23) we are able to express any finite difference formulation of the Overland Flow equations (3-25) and (3-26), or any rearrangement of them, in terms of locally constant quantities of zero and first order of magnitude, and variations in only two dependent variables V and c. We can do this under the mild restriction that the channel is quasi-simple without assuming that the channel is prismatic, although obviously the channel cannot be strongly non-prismatic without violating the assumption in Chapter 2 that all streamlines are approximately parallel.

As discussed in Chapter 4, the locally constant zero order terms cannot contribute to instability over a few time steps, so the locally constant first order terms $\overset{\circ}{f}(s)$ can also be disregarded in our Fourier series representation.

We now introduce our Fourier series representation of the Overland Flow equations term by term. This representation is achieved in five steps as follows:

1. The term is rewritten in its formulation under the difference scheme.

2. The term is separated into zero order locally constant terms, first order locally constant terms, and terms in the variations $\mathring{c}$ and $\mathring{V}$.

3. The term is multiplied by $\Delta t$ to make the coefficients of $\mathring{c}$ and $\mathring{V}$ dimensionless.

4. The variations are expressed as a typical term in a Fourier series such as (A-12) in Appendix A, with the constant term incorporating the locally constant quantities neglected. We follow the notation of Appendix A, writing $\overset{*}{c}^b$, $\overset{*}{V}^b$ for the Fourier transforms of $\mathring{c}_a^b$, $\mathring{V}_a^b$.

5. This term is divided by $e^{ipa\Delta s}$ and expressed as a function of $\theta$, where $\theta = p\Delta s$ as in Appendix A.

As an example of this transformation, we shall transform $\frac{qc}{A}$ through the five steps, using the staggered scheme of Chapter 4. $\frac{qc}{A}$ becomes successively

1. $$\frac{q\frac{1}{2}(c_{a-1}^b + c_{a+1}^b)}{\frac{1}{2}(A_{a-1}^b + A_{a+1}^b)}$$

2. $$\frac{q(c + \frac{1}{2}(\mathring{c}_{a-1}^b + \mathring{c}_{a+1}^b))}{A + \frac{1}{2}(\mathring{A}_{a-1}^b + \mathring{A}_{a+1}^b)} = \frac{qc}{A}\left[1 + \frac{\mathring{c}_{a-1}^b + \mathring{c}_{a+1}^b}{2c} - \frac{\mathring{A}_{a-1}^b + \mathring{A}_{a+1}^b}{2A}\right]$$

$$= \frac{qc}{A}\left[1 + \frac{\mathring{c}_{a-1}^b + \mathring{c}_{a+1}^b}{2c}\left(1 - 2\frac{By}{A}\right)\right] + \mathring{f}(s), \text{ using (5-21).}$$

Note that we obtain the same relation if we introduce $\frac{1}{2}(\mathring{c}_{a-1}^b + \mathring{c}_{a+1}^b)$ directly into (5-21) in place of our typical variation term $\mathring{c}_a^b$, and similarly we can replace $\mathring{c}_a^b$ and $\mathring{V}_a^b$ in

(5-19) to (5-23) by variation terms expressing the particular formulation adopted for any difference scheme.

3. $\frac{qc\Delta t}{A} + \mathring{f}(s)\Delta t + \frac{q}{2A}(1 - 2k)\Delta t\ (\mathring{c}^{b}_{a-1} + \mathring{c}^{b}_{a+1})$

4. $\frac{q}{2A}(1-2k)\Delta t\left[e^{ip(a-1)\Delta s} + e^{ip(a+1)\Delta s}\right] \overset{*}{c}{}^{b}$

5. $\frac{q}{A}\ (1 - 2k)\Delta t\ \cos\theta\ \overset{*}{c}{}^{b}$

We shall denote this complete transformation with an $\Rightarrow$ arrow, writing the above example simply

$$\frac{qc}{A} \Rightarrow \frac{q}{A}\ (1 - 2k)\Delta t\ \cos\theta\ \overset{*}{c}{}^{b}$$

Similarly, in the staggered scheme of Chapter 4

$$V_t \Rightarrow \overset{*}{V}{}^{b+1} - \cos\theta\ \overset{*}{V}{}^{b}$$

For our general one cycle, two level scheme we write

$$V_t \Rightarrow \chi_1 \overset{*}{V}{}^{b+1} - \chi_2 \overset{*}{V}{}^{b} \tag{5-24}$$

where $\chi_1$ and $\chi_2$ are constants or functions of $e^{i\theta}$.

Several mild restrictions are implied by the transformations of the general one cycle, two level scheme which follow. These are that $V_t$ and $w_t$ are formulated in the same manner, and that $V_t$, $w_t$ and the non-homogeneous terms are each formulated in the same way in both characteristic equations (3-25) and (3-26). Because the space differences are often formulated differently along the forward and backward characteristics, we write $V_s$, $w_s$ as $\overset{+}{V}_s$, $\overset{+}{w}_s$ in (3-25) and as $\bar{V}_s$, $\bar{w}_s$ in (3-26). However we still assume that $\overset{+}{V}_s$ and $\overset{+}{w}_s$, $\bar{V}_s$ and $\bar{w}_s$ are formulated in the same manner. Such mild restrictions are necessary for progress in this analysis, but should not exclude many difference schemes.

Using (5-23) we write therefore

$$w_t \Longrightarrow 2k\chi_1 \overset{*}{c}^{b+1} - 2k\chi_2 \overset{*}{c}^{b} \qquad (5\text{-}25)$$

Also

$$\overset{+}{V}_s \Longrightarrow \psi_1 \overset{*}{V}^{b+1} + \psi_2 \overset{*}{V}^{b} \qquad (5\text{-}26)$$

$$\overset{-}{V}_s \Longrightarrow \psi_3 \overset{*}{V}^{b+1} + \psi_4 \overset{*}{V}^{b} \qquad (5\text{-}27)$$

$$\overset{+}{w}_s \Longrightarrow 2k\psi_1 \overset{*}{c}^{b+1} + 2k\psi_2 \overset{*}{c}^{b} \qquad (5\text{-}28)$$

$$\overset{-}{w}_s \Longrightarrow 2k\psi_3 \overset{*}{c}^{b+1} + 2k\psi_4 \overset{*}{c}^{b} \qquad (5\text{-}29)$$

For the non-homogeneous terms we let

$$\frac{gz_s}{\eta} - \frac{gS_f}{\eta} - \frac{qU}{A\eta} \Longrightarrow \tau_1\Delta t \ \overset{*}{c}^{b+1} + \tau_2\Delta t \ \overset{*}{c}^{b} + \tau_3\Delta t \ \overset{*}{V}^{b+1} + \tau_4\Delta t \ \overset{*}{V}^{b} \qquad (5\text{-}30)$$

$$(q - VH_s)\frac{c}{A} \Longrightarrow \upsilon_1\Delta t \ \overset{*}{c}^{b+1} + \upsilon_2\Delta t \ \overset{*}{c}^{b} + \upsilon_3\Delta t \ \overset{*}{V}^{b+1} + \upsilon_4\Delta t \ \overset{*}{V}^{b} \qquad (5\text{-}31)$$

We have written the coefficients $\tau\Delta t$ and $\upsilon\Delta t$ to indicate that they are of $O(\Delta t)$ in contrast to $\chi$ and $\psi$.

Thus the transformations of (3-25) and (3-26) are respectively

$$(2k\chi_1+(V+c)2k\psi_1-\tau_1\Delta t-\upsilon_1\Delta t)\overset{*}{c}^{b+1} + (\chi_1+(V+c)\psi_1-\tau_3\Delta t-\upsilon_3\Delta t)\overset{*}{V}^{b+1}$$

$$= (2k\chi_2-(V+c)2k\psi_2+\tau_2\Delta t+\upsilon_2\Delta t)\overset{*}{c}^{b} + (\chi_2-(V+c)\psi_2+\tau_4\Delta t+\upsilon_4\Delta t)\overset{*}{V}^{b} \qquad (5\text{-}32)$$

$$(-2k\chi_1-(V-c)2k\psi_3-\tau_1\Delta t+\upsilon_1\Delta t)\overset{*}{c}^{b+1} + (\chi_1+(V-c)\psi_3-\tau_3\Delta t+\upsilon_3\Delta t)\overset{*}{V}^{b+1}$$

$$= (-2k\chi_2+(V-c)2k\psi_4+\tau_2\Delta t-\upsilon_2\Delta t)\overset{*}{c}^{b} + (\chi_2-(V-c)\psi_4+\tau_4\Delta t-\upsilon_4\Delta t)\overset{*}{V}^{b} \qquad (5\text{-}33)$$

Note that variations in the coefficients V+c, V-c contribute only second order quantities, so these coefficients are locally constant.

It is convenient to subtract, then add (5-32) and (5-33) as this recasts the equations in the form usually encountered in explicit finite difference schemes, and facilitates comparison with special stability analyses, for instance that of Chapter 4. However this does not bar the application of this general analysis to implicit schemes.

Subtracting (5-33) from (5-32) and dividing by two gives

$$(2k\Psi_1 - \nu_1\Delta t)\overset{*}{c}^{b+1} + (\Phi_1 - \nu_3\Delta t)\overset{*}{V}^{b+1} = (2k\Psi_2 + \nu_2\Delta t)\overset{*}{c}^{b} + (\Phi_2 + \nu_4\Delta t)\overset{*}{V}^{b} \tag{5-34}$$

Adding (5-33) to (5-32) and dividing by two gives

$$(2k\Phi_1 - \tau_1\Delta t)\overset{*}{c}^{b+1} + (\Psi_1 - \tau_3\Delta t)\overset{*}{V}^{b+1} = (2k\Phi_2 + \tau_2\Delta t)\overset{*}{c}^{b} + (\Psi_2 + \tau_4\Delta t)\overset{*}{V}^{b} \tag{5-35}$$

where

$$\Psi_1 = \chi_1 + \tfrac{1}{2}\left[(V + c)\psi_1 + (V - c)\psi_3\right] \tag{5-36}$$

$$\Psi_2 = \chi_2 - \tfrac{1}{2}\left[(V + c)\psi_2 + (V - c)\psi_4\right] \tag{5-37}$$

$$\Phi_1 = \tfrac{1}{2}\left[(V + c)\psi_1 - (V - c)\psi_3\right] \tag{5-38}$$

$$\Phi_2 = \tfrac{1}{2}\left[-(V + c)\psi_2 + (V - c)\psi_4\right] \tag{5-39}$$

## 5.7 Eigenvalues of the Amplification Matrix

If we write $M_1 = \begin{bmatrix} \Upsilon_1 & \Upsilon_2 \\ \Upsilon_3 & \Upsilon_4 \end{bmatrix}$ , $M_2 = \begin{bmatrix} \Upsilon_5 & \Upsilon_6 \\ \Upsilon_7 & \Upsilon_8 \end{bmatrix}$

where $\Upsilon_1 = 2k\Psi_1 - \nu_1\Delta t$ etc., we can write (5-34) and (5-35)

$$M_1 \begin{bmatrix} \overset{*}{c}^{b+1} \\ \overset{*}{V}^{b+1} \end{bmatrix} = M_2 \begin{bmatrix} \overset{*}{c}^{b} \\ \overset{*}{V}^{b} \end{bmatrix}$$

or, if det $(M_1) \neq 0$, $$\begin{bmatrix} \overset{*}{c}^{b+1} \\ \overset{*}{V}^{b+1} \end{bmatrix} = M_1^{-1} M_2 \begin{bmatrix} \overset{*}{c}^{b} \\ \overset{*}{V}^{b} \end{bmatrix} \tag{5-40}$$

Note that in Chapter 5 and 6 we shall write det $(M_1)$ for the determinant of a matrix $M_1$ rather than the common $|M_1|$ in order to avoid confusion with our notation $|\lambda|$ for the modulus of a scalar $\lambda$.

Now $$M_1^{-1} M_2 = \frac{1}{\det(M_1)} \begin{bmatrix} \Upsilon_4 \Upsilon_5 - \Upsilon_2 \Upsilon_7 & \Upsilon_4 \Upsilon_6 - \Upsilon_2 \Upsilon_8 \\ -\Upsilon_3 \Upsilon_5 + \Upsilon_1 \Upsilon_7 & -\Upsilon_3 \Upsilon_6 + \Upsilon_1 \Upsilon_8 \end{bmatrix}$$

$$= \frac{M_3}{\det(M_1)} \quad \text{say} \tag{5-41}$$

The amplification matrix $M_a = M_1^{-1} M_2$ has eigenvalues $\lambda_{1,2}$ given by $\det(M_a - \lambda I) = 0$. Now

$$\lambda_{1,2} = \frac{1}{\det(M_1)} \Lambda_{1,2} \tag{5-42}$$

where $\Lambda_{1,2}$ are the eigenvalues of $M_3$.

If we write

$$M_3 = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} \tag{5-43}$$

then $$\Lambda_{1,2} = \tfrac{1}{2}(\mu_{11} + \mu_{22}) \pm \sqrt{\Omega} \tag{5-44}$$

where $$\Omega = \tfrac{1}{4}(\mu_{11} - \mu_{22})^2 + \mu_{21}\mu_{12} \tag{5-45}$$

and

$$\mu_{11} = 2k(\Psi_1\Psi_2-\Phi_1\Phi_2) + \Delta t(\nu_2\Psi_1-2k\tau_3\Psi_2-\tau_2\Phi_1+2k\nu_3\Phi_2) + \Delta t^2(\nu_3\tau_2-\nu_2\tau_3) \tag{5-46}$$

$$\mu_{22} = 2k(\Psi_1\Psi_2-\Phi_1\Phi_2) + \Delta t(2k\tau_4\Psi_1-\nu_1\Psi_2-2k\nu_4\Phi_1+\tau_1\Phi_2) + \Delta t^2(\nu_4\tau_1-\nu_1\tau_4) \tag{5-47}$$

$$\mu_{12} = \Psi_1\Phi_2-\Phi_1\Psi_2 + \Delta t(\nu_4\Psi_1+\nu_3\Psi_2-\tau_4\Phi_1-\tau_3\Phi_2) + \Delta t^2(\nu_3\tau_4-\nu_4\tau_3) \tag{5-48}$$

$$\mu_{21} = 4k^2(\Psi_1\Phi_2-\Phi_1\Psi_2) + 2k\Delta t(\tau_2\Psi_1+\tau_1\Psi_2-\nu_2\Phi_1-\nu_1\Phi_2) + \Delta t^2(\nu_2\tau_1-\nu_1\tau_2) \tag{5-49}$$

Also we have

$$\det(M_1) = 2k(\Psi_1^2-\Phi_1^2) - \Psi_1\Delta t(\nu_1+2k\tau_3) + \Phi_1\Delta t(\tau_1+2k\nu_3) + \Delta t^2(\nu_1\tau_3-\nu_3\tau_1) \tag{5-50}$$

## 5.8 Stability as Mesh Lengths Tend to Zero

We can assume that $r = \Delta t/\Delta s$ remains constant as $\Delta t, \Delta s$ tend to zero and thus simply neglect all terms including the factor $\Delta t$ in this limiting case.

From equations (5-45) to (5-49) in this case

$$\Omega = 4k^2(\Psi_1\Phi_2 - \Phi_1\Psi_2)^2$$

Thus from (5-44), (5-46) and (5-47)

$$\Lambda = 2k\left[\Psi_1\Psi_2 - \Phi_1\Phi_2 \pm (\Psi_1\Phi_2 - \Phi_1\Psi_2)\right] \tag{5-51}$$

From (5-50) $\det(M_1) = 2k(\Psi_1^2 - \Phi_1^2)$

Hence from (5-42) and (5-51)

$$\lambda_1 = \frac{\Psi_2 + \Phi_2}{\Psi_1 + \Phi_1} = \frac{\chi_2 - (V+c)\psi_2}{\chi_1 + (V+c)\psi_1} \tag{5-52}$$

using (5-36), (5-37), (5-38), (5-39). Similarly

$$\lambda_2 = \frac{\chi_2 - (V-c)\psi_4}{\chi_1 + (V-c)\psi_3} \tag{5-53}$$

While it is the purpose of this thesis to deal with mesh lengths of practical size, we must require any valid difference solution to be convergent. Thus a necessary condition for the validity of any difference solution is its stability in the limit as the mesh lengths tend to zero. We present two examples of the application of this condition.

For the CIR rectangular scheme examined in Appendix A, from (A-3) and (5-24), (5-25), $\chi_1 = \chi_2 = 1$. From (A-4) and (5-26), (5-28), $\psi_1 = 0$, $\psi_2 = r(1 - e^{-ip\Delta s}) = r\ (1 - \cos\theta + i \sin\theta)$. From (A-5) and (5-27), (5-29), $\psi_3 = 0$,

$$\psi_4 = r(e^{ip\Delta s} - 1) = r(\cos\theta - 1 + i \sin\theta).$$

Thus equations (5-52) and (5-53) combine to give (A-22) as expected.

For the implicit scheme discussed by Liggett and Woolhiser (1967) $\chi_1 = \chi_2 = 1$, $\psi_1 = \psi_3 = \psi_2 = \psi_4 = \frac{r}{4}(e^{ip\Delta s} - e^{-ip\Delta s}) = \frac{ir}{2} \sin\theta$

From (5-52)

$$\lambda_1 = \frac{(1 - \frac{ir}{2}(V+c)\ \sin\theta)(1 - \frac{ir}{2}(V+c)\ \sin\theta)}{(1 + \frac{ir}{2}(V+c)\ \sin\theta)(1 - \frac{ir}{2}(V+c)\ \sin\theta)}$$

$$= \frac{1 - \frac{r^2}{4}\ (V+c)^2\ \sin^2\theta\ - ir(V+c)\ \sin\theta}{1 + \frac{r^2}{4}\ (V+c)^2\ \sin^2\theta}$$

$$|\lambda_1|^2 = \frac{1 - \frac{r^2}{2}(V+c)^2\sin^2\theta + \frac{r^4}{16}(V+c)^4\sin^4\theta + r^2(V+c)^2\sin^2\theta}{(1 + \frac{r^2}{4}(V+c)^2\ \sin^2\theta)^2}$$

$$= 1$$

We similarly find from (5-53) that $|\lambda_2|^2 = 1$ which is independent of r, V, c and $\theta$ and hence this implicit formulation is unconditionally stable, from (4-56), as the mesh lengths tend to zero.

## 5.9 General Criteria for Physical Stability

We now consider stability as $\theta$ tends to zero, because as we remarked in Section 4.7, the eigenvalues of any function in real constants and $e^{i\theta}$ will have turning points with respect to variation in $\theta$ when $\sin\theta = 0$, and hence there is possibly a maximum of $|\lambda|$ when $\theta$ is zero. The case $\theta = 0$ is therefore very important, and can be studied at this point because $\chi_1$, $\chi_2$, $\psi_1$, $\psi_2$, $\psi_3$ and $\psi_4$ are independent of the difference scheme when $\theta$ tends to zero. This follows because if $\theta = 0$ we are studying a harmonic with a period which fits an integral number of times into the length $\Delta s$, and such harmonics represent that part of the variation quantity which is constant at each mesh point on a time line. Thus any average over the mesh points simply gives the constant value at any mesh point, while any difference between values at mesh points is zero. Hence for any reasonable difference scheme, $\chi_1 = \chi_2 = 1$ when $\theta = 0$, while $\psi_{1,2,3,4} = 0$.

We postpone the case $\theta = 0$ to Section 5.10, considering in this section the case where $\theta$ tends to zero. We therefore assume that $\theta$ is small such that terms $O(\theta^2)$ can be dropped.

Thus $\quad e^{i\theta} \;=\; 1 + i\theta$

and hence we can say, from (5-36), (5-37), (5-38) and (5-39)

$$\Psi_1 = 1 + in_1\theta$$

$$\Psi_2 = 1 + in_2\theta$$

$$\Phi_1 = in_3\theta$$

$$\Phi_2 = in_4\theta$$

where $n_1$, $n_2$, $n_3$ and $n_4$ are real constants. These equations hold because $\chi_1$ and $\chi_2$ are averages of terms in $e^{i\theta}$, $e^{-i\theta}$ etc, while $\psi_{1,2,3,4}$ are differences between terms in $e^{i\theta}$, $e^{-i\theta}$ etc.

Now in order to evaluate $\Omega$ from (5-45) in this general case we take $\Delta t$ as small, so that we can drop from $\Omega$ all terms $O(\theta^3, \theta^2\Delta t, \theta\Delta t^2, \Delta t^3)$ and all higher order terms. Thus using (5-45), (5-46) and (5-47)

$$\tfrac{1}{4}(\mu_{11} - \mu_{22})^2 = \frac{\Delta t^2}{4}\,(\nu_2 + \nu_1 + 2k(\tau_4 + \tau_3))^2 \tag{5-54}$$

Using (5-45), (5-48) and (5-49)

$$\mu_{12}\mu_{21} = -4k^2(n_4-n_3)^2\theta^2 + i2k(n_4-n_3)\theta\Delta t\left[\tau_2+\tau_1+2k(\nu_4+\nu_3)\right] + 2k\Delta t^2(\tau_2+\tau_1)(\nu_4+\nu_3) \tag{5-55}$$

$$\begin{aligned}\Omega = {} & -4k^2(n_4-n_3)^2\theta^2 + i2k(n_4-n_3)\theta\Delta t\left[\tau_2+\tau_1+2k(\nu_4+\nu_3)\right] \\ & + \frac{\Delta t^2}{4}\left[\tau_2+\tau_1+2k(\nu_4+\nu_3)\right]^2 - \frac{\Delta t^2}{4}\left[\tau_2+\tau_1-2k(\nu_4+\nu_3)\right]^2 \\ & + \frac{\Delta t^2}{4}\left[\nu_2+\nu_1-2k(\tau_4+\tau_3)\right]^2\end{aligned} \tag{5-56}$$

If we neglect the last two terms, $\Omega$ becomes a perfect square and $\sqrt{\Omega}$ follows immediately. We achieve this simplification by taking terms $O(\Delta t^2)$ to be small compared with those of $O(\theta^2)$, giving

$$\sqrt{\Omega} = \left[\tau_2+\tau_1+2k(\nu_4+\nu_3)\right]\frac{\Delta t}{2} + i2k(n_4-n_3)\theta \tag{5-57}$$

From (5-46) and (5-47), dropping terms of $O(\theta^2,\theta\Delta t,\Delta t^2)$ and higher orders

$$\tfrac{1}{2}(\mu_{11}+\mu_{22}) = 2k\left[1+i(n_1+n_2)\theta\right] + \frac{\Delta t}{2}\left[\nu_2-\nu_1+2k(\tau_4-\tau_3)\right] \tag{5-58}$$

Using (5-44), (5-57) and (5-58)

$$\begin{aligned}\Lambda &= 2k + \frac{\Delta t}{2}\left[\nu_2-\nu_1+2k(\tau_4-\tau_3) \pm\left[\tau_2+\tau_1+2k(\nu_4+\nu_3)\right]\right] \\ &\quad + i2k\theta\left[(n_1+n_2) \pm (n_4-n_3)\right]\end{aligned} \tag{5-59}$$

From (5-50), dropping terms of $O(\theta^2,\theta\Delta t,\Delta t^2)$ and higher orders

$$\det(M_1) = 2k - \Delta t(\nu_1 + 2k\tau_3) + i4kn_1\theta$$

$$\begin{aligned}\frac{1}{\det(M_1)} &= \frac{1}{2k}\left(1 - \frac{\Delta t}{2k}(\nu_1 + 2k\tau_3) + i2n_1\theta\right)^{-1} \\ &= \frac{1}{2k}\left(1 + \frac{\Delta t}{2k}(\nu_1 + 2k\tau_3) - i2n_1\theta\right)\end{aligned} \tag{5-60}$$

From (5-42) we get $\lambda$ by multiplying $\Lambda$ by $1/\det(M_1)$, and after dropping terms of $O(\theta^2,\theta\Delta t,\Delta t^2)$ it is clear that all terms of $O(\theta)$ are imaginary, while no terms of $O(1)$ are imaginary. Therefore when we take the magnitude $|\lambda|^2$ the imaginary contribution is of $O(\theta^2)$ and can be dropped. We have assumed terms of $O(\theta^2)$ are large compared with terms of $O(\Delta t^2)$ so these must also be dropped from $|\lambda|^2$, but we can choose $\Delta t$ and $\theta$ such that terms $O(\Delta t)$ are large compared with terms $O(\theta^2)$, and in this case it is necessary to retain terms $O(\Delta t)$ in $|\lambda|^2$. For instance, we can choose $\theta$ and $\Delta t$ such that terms $O(\theta)$ are about $10^{-2}$ and terms $O(\Delta t)$ are about $10^{-3}$.

After the above manipulation, we find

$$|\lambda| = 1 + \frac{\Delta t}{4k}\left[\upsilon_2+\upsilon_1+2k(\tau_4+\tau_3) \pm \left[\tau_2+\tau_1 + 2k(\upsilon_4+\upsilon_3)\right]\right] \qquad (5\text{-}61)$$

Thus from (4-56) we have two stability criteria which depend only on the locally constant flow parameters $\tau$ and $\upsilon$, and these can be expressed

$$\upsilon_2 + \upsilon_1 + 2k(\tau_4 + \tau_3) \leq 0 \qquad (5\text{-}62)$$

$$|\tau_2 + \tau_1 + 2k(\upsilon_4+\upsilon_3)| \leq -\upsilon_2 - \upsilon_1 - 2k(\tau_4+\tau_3) \qquad (5\text{-}63)$$

These stability criteria have been proved for $\Delta t$ small and both must hold as $\Delta t$ tends to zero as we can always define a suitable $\theta$ for the above proof to hold. Thus these criteria are associated with instability of the flow described by the exact partial differential equations, which must imply that the flow is physically unstable if either of the above criteria are infringed. We shall show in Section 6.3 that the above general criteria, when applied to the special case of a prismatic channel without lateral flows, reduce to the existing criteria discussed in Section 5.4.

## 5.10 General Stability Criteria for Near Prismatic Channels

We now examine the case $\theta = 0$, where we expect to find a turning point in the value of $|\lambda|$.

Now $\Psi_1 = \Psi_2 = 1$ and $\Phi_1 = \Phi_2 = 0$ for all formulations.

Examination of (5-48) reveals that if $\upsilon_3 = \upsilon_4 = 0$, then $\mu_{12} = 0$ in this case. Now from (5-31) $\upsilon_3$ and $\upsilon_4$ arise from the term $-VH_s c/A$, and if $H_s$ is small, that is, the channel is near prismatic, we can regard $\upsilon_3$ and $\upsilon_4$ as negligible* and therefore set $\mu_{12} = 0$ when $\theta = 0$. Thus from (5-44) and (5-45)

$$\Lambda = \tfrac{1}{2}(\mu_{11} + \mu_{22}) \pm \tfrac{1}{2}(\mu_{11} - \mu_{22})$$

$$= \mu_{11} \text{ or } \mu_{22} \tag{5-64}$$

From (5-46), (5-47) and (5-50)

$$\mu_{11} = 2k + \Delta t(\upsilon_2 - 2k\tau_3) - \Delta t^2 \upsilon_2 \tau_3 = (2k + \upsilon_2 \Delta t)(1 - \tau_3 \Delta t) \tag{5-65}$$

$$\mu_{22} = 2k + \Delta t(2k\tau_4 - \upsilon_1) - \Delta t^2 \upsilon_1 \tau_4 = (2k - \upsilon_1 \Delta t)(1 + \tau_4 \Delta t) \tag{5-66}$$

$$\det(M_1) = 2k - \Delta t(\upsilon_1 + 2k\tau_3) + \Delta t^2 \upsilon_1 \tau_3 = (2k - \upsilon_1 \Delta t)(1 - \tau_3 \Delta t) \tag{5-67}$$

Thus from (5-42)

$$\lambda_1 = \frac{2k + \upsilon_2 \Delta t}{2k - \upsilon_1 \Delta t} \tag{5-68}$$

$$\lambda_2 = \frac{1 + \tau_4 \Delta t}{1 - \tau_3 \Delta t} \tag{5-69}$$

By applying (4-56) we have two more stability criteria, both of which must be satisfied:

$$\left| \frac{2k + \upsilon_2 \Delta t}{2k - \upsilon_1 \Delta t} \right| \leqslant 1 \tag{5-70}$$

$$\left| \frac{1 + \tau_4 \Delta t}{1 - \tau_3 \Delta t} \right| \leqslant 1 \tag{5-71}$$

Note that we have not restricted $\Delta t$ in any way in

---

*Note that if $\tau_1$ and $\tau_2$ are negligible, we can set $\mu_{21} = 0$ when $\theta = 0$ and (5-64) again follows from (5-44) and (5-45). See also Section 6.5.

deriving these criteria, so that (5-70) and (5-71) hold for all $\Delta t$ in near prismatic channels as long as the basic assumptions in Section 5.2 hold. We shall restate these criteria more explicitly in terms of flow properties in Section 6.4.

# CHAPTER 6

## COMPLETE STABILITY ANALYSIS ON SELECTED FINITE DIFFERENCE SCHEMES

### 6.1 Evaluation of the Non-Homogeneous Terms

We must be able to relate $\tau_{1,2,3,4}$ and $\nu_{1,2,3,4}$ of equations (5-30), (5-31) to locally constant properties of the flow solution, as well as to the space dependence of the method used by the difference scheme, in order to evaluate the non-homogeneous terms explicitly. We shall therefore assume in this chapter that the difference scheme is formulated in such a way that a single value of c, and of V, is used in the evaluation of all non-homogeneous terms throughout each individual application of the difference scheme. This is a common but not universal practice in finite difference schemes (see Terzidis and Strelkoff (1965)*), but no good reason for its abandonment has been advanced. We denote these values $c + \mathring{c}^{b+1} + \mathring{c}^{b}$, $V + \mathring{V}^{b+1} + \mathring{V}^{b}$, and also assume that both are obtained by the same averaging process applied to the respective solutions at mesh points on the time levels $b\Delta t$ and $(b+1)\Delta t$. Thus we can say that $\mathring{c}^{b+1}$ transforms to $\alpha_1\overset{*}{c}^{b+1}$, $\mathring{c}^{b}$ to $\alpha_2\overset{*}{c}^{b}$, $\mathring{V}^{b+1}$ to $\alpha_1\overset{*}{V}^{b+1}$, and $\mathring{V}^{b}$ to $\alpha_2\overset{*}{V}^{b}$, where $\alpha_1$ and $\alpha_2$ are constants or functions of $\theta$ representing the space dependence of the averaging process on the two levels. We expect $\alpha_1 + \alpha_2$ to be unity when

---

*Terzidis and Strelkoff cite schemes which evaluate the friction term and lateral flow terms at different points in the solution, although they themselves recommend the use of a single value of c and of V in all non-homogeneous terms.

$\theta = 0$, as otherwise $c + \overset{\circ}{c}^{b+1} + \overset{\circ}{c}^{b}$ and $V + \overset{\circ}{V}^{b+1} + \overset{\circ}{V}^{b}$ could not be average values of c and V (see Section 5.9).

For an example, we refer to the staggered scheme discussed in Chapter 4 and again in Section 5.6.

In this case

$$c + \overset{\circ}{c}^{b+1} + \overset{\circ}{c}^{b} = \tfrac{1}{2}(c^{b}_{a-1} + c^{b}_{a+1}) = c + \tfrac{1}{2}(\overset{\circ}{c}^{b}_{a-1} + \overset{\circ}{c}^{b}_{a+1})$$

Therefore

$$\overset{\circ}{c}^{b+1} = 0 \qquad \overset{\circ}{c}^{b} = \tfrac{1}{2}(\overset{\circ}{c}^{b}_{a-1} + \overset{\circ}{c}^{b}_{a+1})$$

Similarly

$$\overset{\circ}{V}^{b+1} = 0 \qquad \overset{\circ}{V}^{b} = \tfrac{1}{2}(\overset{\circ}{V}^{b}_{a-1} + \overset{\circ}{V}^{b}_{a+1})$$

Thus

$$\alpha_1 = 0 \qquad \alpha_2 = \tfrac{1}{2}(e^{-ip\Delta s} + e^{ip\Delta s}) = \cos\theta$$

In Section 5.6 we showed the full transformation of the term $\frac{qc}{A}$ for the staggered scheme. The term was expanded to

$$\frac{qc}{A}\left[1 + \frac{\overset{\circ}{c}^{b}_{a-1} + \overset{\circ}{c}^{b}_{a+1}}{2c}(1-2k)\right] + \overset{\circ}{f}(s) = \frac{qc}{A}\left[1 + \frac{\overset{\circ}{c}^{b+1} + \overset{\circ}{c}^{b}}{c}(1-2k)\right] + \overset{\circ}{f}(s) \tag{6-1}$$

We can also, by the same argument, replace $\overset{\circ}{c}^{b}_{a}$ in (5-21) by $\overset{\circ}{c}^{b+1} + \overset{\circ}{c}^{b}$ whatever averaging process these variations represent. Thus we have the general transformation, following the five steps defined in Section 5.6

$$\frac{qc}{A} \Longrightarrow \frac{q}{A}(1-2k)\Delta t(\alpha_1 \overset{*}{c}^{b+1} + \alpha_2 \overset{*}{c}^{b})$$

Similarly, using (5-21) and (5-22)

$$(q-VH_s)\frac{c}{A} \Longrightarrow \left(\frac{q}{A} - \frac{VH_s}{A}\right)(1-2k)\Delta t(\alpha_1 \overset{*}{c}^{b+1} + \alpha_2 \overset{*}{c}^{b}) - \frac{cH_s\Delta t}{A}(\alpha_1 \overset{*}{V}^{b+1} + \alpha_2 \overset{*}{V}^{b}) \tag{6-2}$$

In the same way, using (5-19) and (5-20)

$$\frac{qz_s}{\eta} - \frac{gS_f}{\eta} - \frac{qU}{A\eta} \Longrightarrow \left[\frac{gS_f}{\eta}\,\frac{2m}{jc} + \frac{2kqU}{A\eta c}\right]\Delta t(\alpha_1 \overset{*}{c}^{b+1} + \alpha_2 \overset{*}{c}^{b}) - \left[\frac{gS_f}{j\eta V} + \frac{q}{\eta A}\right]\Delta t(\alpha_1 \overset{*}{V}^{b+1} + \alpha_2 \overset{*}{V}^{b}) \tag{6-3}$$

Define

$$\sigma = \frac{gS_f}{2\eta jV} \tag{6-4}$$

$$\gamma_1 = \frac{q}{2A\eta} \tag{6-5}$$

$$\gamma_2 = \frac{q}{2A} \tag{6-6}$$

where $\sigma$ and $\gamma_1$, $\gamma_2$ incorporate the $\sigma$ and $\gamma$ defined in (4-33) for the special case of Chapter 4. Note that $\sigma$ is always positive because $S_f$ takes the same sign as V. We distinguish between $\gamma_1$ and $\gamma_2$ even though $\eta$ is commonly unity because $\gamma_1$ will be associated with the momentum inflow term, while $\gamma_2$ will be associated simply with volume inflow. Note that $\gamma_1$ and $\gamma_2$ both take the same sign as q.

We introduce

$$G_1 = 4(mF\sigma + k\gamma_1\frac{U}{c}) \tag{6-7}$$

$$G_2 = 2(\sigma + \gamma_1) \tag{6-8}$$

$$G_3 = (2k-1)(2\gamma_2 - \frac{VH_s}{A}) \tag{6-9}$$

$$G_4 = \frac{cH_s}{A} \tag{6-10}$$

Thus, from (6-3) and (5-30)

$$\tau_1 = G_1\alpha_1 \qquad \tau_2 = G_1\alpha_2 \qquad \tau_3 = -G_2\alpha_2 \qquad \tau_4 = -G_2\alpha_2 \tag{6-11}$$

From (6-2) and (5-31)

$$\upsilon_1 = -G_3\alpha_1 \qquad \upsilon_2 = -G_3\alpha_2 \qquad \upsilon_3 = -G_4\alpha_1 \qquad \upsilon_4 = -G_4\alpha_2 \qquad (6\text{-}12)$$

## 6.2 Relative Magnitudes of the Coefficients

The relative magnitudes of $G_1$, $G_2$, $G_3$ and $G_4$ can readily be assessed from (6-7) to (6-10) if the relative magnitudes of $\sigma$, $\gamma_1$, $\gamma_2$ and $G_4$ are known, assuming that m, F, k and U/c can be estimated for individual cases. $\eta$ should similarly be available, so we need only compare $\sigma$, $\gamma_1$ and $G_4$, which express the importance in the stability analysis of, respectively, the frictional resistance, lateral flows, and channel shape and width variations.

From (6-4) and (6-5)

$$\frac{\gamma_1}{\sigma} = \frac{qjV}{AgS_f} \qquad (6\text{-}13)$$

For laminar flow, j = 1 and, from (2-52)

$$S_f = \frac{V\mu(k+2)}{\rho g y^2}$$

$$\frac{\gamma_1}{\sigma} = \frac{q\rho y^2}{A\mu(k+2)} = \frac{q}{B}\,\frac{k}{k+2}\,\frac{\rho y}{\mu} \qquad (6\text{-}14)$$

Now q/B is the lateral inflow (outflow if negative) per unit surface area, which is conveniently regarded as the average rate of rise of the surface of a body of water with no outlet. This is the normal measure, for instance, of rainfall rate. For water $\rho/\mu$ = 100 sec/cm$^2$ approx and it is difficult to imagine laminar flow deeper than 0.5 cm. Taking $k_{max}$ = 2, we get

$$\frac{\gamma_1}{\sigma} < 25 \frac{\text{sec}}{\text{cm}} \frac{q}{B} \tag{6-15}$$

A maximum rainfall rate might be $4 \times 10^{-3}$ cm/sec = 6 in/hr, in which case $\gamma_1/\sigma < 0.1$, so we can assume that for laminar flow $\gamma_1$ is an order of magnitude smaller than $\sigma$.

For turbulent flow, $j = \frac{1}{2}$ and, from (2-40) and (2-41)

$$S_f = \frac{V^2 n^2}{M^2 R^{4/3}}$$

From (6-13)

$$\frac{\gamma_1}{\sigma} = \frac{q\ M^2 R}{2AVn^2 g} = \frac{qR}{Q} \frac{M^2 R^{\frac{1}{3}}}{2n^2 g} \tag{6-16}$$

$M^2 R^{\frac{1}{3}}/2n^2 g$ will generally not exceed 100, although for power canals with very smooth lining, or for very large rivers, a value of 200-300 might be reached. Thus if the lateral inflow along a length of about 1000R of the channel is less than Q, $\gamma_1$ can be regarded as an order of magnitude smaller than $\sigma$. Otherwise $\gamma_1$ and $\sigma$ are of the same order unless flow along the channel is negligible.

From (6-10) and (6-4)

$$\frac{G_4}{\sigma} = \frac{2\eta j}{S_f} \frac{cVH_s}{gA} \tag{6-17}$$

For laminar flow, using $j = 1$ and (2-52) again

$$\frac{G_4}{\sigma} = \frac{2\eta\rho c y^2 H_s}{\mu(k+2)A} \tag{6-18}$$

Using (3-14) and (2-35) to evaluate c

$$\frac{G_4}{\sigma} = \frac{2}{k+2}\sqrt{\frac{\eta \cos\phi}{k}}\ \frac{\rho g^{\frac{1}{2}} y^{\frac{3}{2}}}{\mu} \frac{yH_s}{A} \tag{6-19}$$

For y = 0.5 cm, $\frac{\rho g^{\frac{1}{2}} y^{\frac{3}{2}}}{\mu}$ = 1100 approx - thus we can say, taking minimum k = 1, max $\eta \cos\phi \approx 1$

$$\frac{G_4}{\sigma} < 1000 \frac{yH_s}{A} \tag{6-20}$$

Thus if the channel is such that it does not double in area, measured at constant depth, over 50 metres (0.5 cm x $10^4$)

$$\frac{G_4}{\sigma} < 0.1 \tag{6-21}$$

For turbulent flow, using j = ½ and (2-40) and (2-41) again, from (6-17)

$$\frac{G_4}{\sigma} = \frac{\eta}{F} \frac{M^2 R^{\frac{1}{3}}}{n^2 g} \frac{RH_s}{A} \tag{6-22}$$

Because $\frac{M^2 R^{\frac{1}{3}}}{n^2 g} < 300$ and $\eta \approx 1$, we have

$$\frac{FG_4}{\sigma} < 300 \frac{RH_s}{A} \tag{6-23}$$

Thus if the channel is such that it does not double in area, measured at constant depth, over 3000R,

$$\frac{FG_4}{\sigma} < 0.1 \tag{6-24}$$

Note that common values of $\frac{M^2 R^{\frac{1}{3}}}{n^2 g}$ are about 100 for rivers, in which case (6-24) holds if the channel does not double in area, measured at a constant depth, over 1000R.

## 6.3 Interpretation of the General Physical Stability Criteria

We are now able to interpret the stability criteria (5-62) and (5-63). Note that they are not independent, as (5-63) implicitly requires (5-62) to hold. Thus we need consider only (5-63) which becomes, using (6-11) and (6-12)

$$\left|G_1(\alpha_1+\alpha_2) - 2kG_4(\alpha_1+\alpha_2)\right| \leqslant G_3(\alpha_1+\alpha_2) + 2kG_2(\alpha_1+\alpha_2) \qquad (6\text{-}25)$$

Now, for $\theta = 0$, $\alpha_1 + \alpha_2 = 1$ and hence for $\theta$ small $\alpha_1 + \alpha_2$ is a positive number and can be cancelled throughout, so that (6-25) becomes

$$\left|G_1 - 2kG_4\right| \leqslant G_3 + 2kG_2 \qquad (6\text{-}26)$$

Thus, as might be expected of physical stability criteria, (5-63) is independent of the difference formulation as well as $\Delta t$, because (6-26) contains only physical flow parameters. In fact (6-26) is independent of the assumption made in Section 6.1 that the non-homogeneous terms are all evaluated using a single value of c and of V at each application of the difference scheme, because the weights assigned to any selection of mesh points must sum to unity for the weighted sum to be in any sense an average, and it can now be seen that when $\theta = 0$ any averaging process must therefore give $\tau_2 + \tau_1 = G_1$, $\tau_4 + \tau_3 = -G_2$, $\vartheta_2 + \vartheta_1 = -G_3$ and $\vartheta_4 + \vartheta_3 = -G_4$.

Using (6-7) to (6-10) in (6-26) we have

$$\left|4(mF\sigma + k\gamma_1 \frac{U}{c}) - 2kG_4\right| \leqslant 4k(\sigma + \gamma_1) + (2k-1)(2\gamma_2 - FG_4) \qquad (6\text{-}27)$$

Because $\sigma$ is always positive (ignoring the case of zero flow), we may divide (6-27) throughout by $\sigma$ to express (5-63) as

$$\left|4(mF + k\frac{\gamma_1 U}{\sigma c}) - 2k\frac{G_4}{\sigma}\right| \leqslant 4k(1 + \frac{\gamma_1}{\sigma}) + (2k-1)(2\frac{\gamma_2}{\sigma} - F\frac{G_4}{\sigma}) \qquad (6\text{-}28)$$

A number of criteria are incorporated in (6-28), depending on the relative magnitudes of the individual parameters. Now m lies between about $\frac{1}{2}$ and 2, while k varies from 1 to 2 or at most 3, so that both m and k are always O(1). We can also assume that $\gamma_2$ is of $O(\gamma_1)$ throughout, so that we need consider wide variations in only four dimensionless parameters, F, $\gamma_1/\sigma$, U/c and $G_4/\sigma$. U/c is secondary, requiring consideration only if $\gamma_1/\sigma$ is appreciable.

We therefore treat (6-28) in four cases, depending on the relative magnitudes of $|F|$, $|G_4|/\sigma$, and $|\gamma_1|/\sigma$. Note that the magnitude of the frictional resistance parameter $\sigma$ is closely related to $|F|$, as the larger the Froude Number of a given discharge, the greater the velocity and smaller the depth, and hence from (2-53) the greater the friction slope. $|\gamma_1|$ and $|G_4|$ also increase with $|F|$ when the depth decreases, but to a lesser extent than $\sigma$. Before applying (6-28) to a particular problem, an estimate of F must be made, with which $|G_4|/\sigma$ can be evaluated for turbulent flow from (6-22). For laminar flow $|G_4|/\sigma$ is given by (6-19) and $|\gamma_1|/\sigma$ by (6-14), while for turbulent flow $|\gamma_1|/\sigma$ is given by (6-16).

Case 1 $|F|$ appreciable, $|G_4|/\sigma$ negligible, $|\gamma_1|/\sigma$ negligible, i.e. an appreciable flow through a channel with little shape or width variation (a near prismatic channel) and with negligible lateral flow. Note that $FG_4/\sigma$ is taken as negligible compared with mF.

(6-28) becomes

$$|mF| \leqslant k$$

which is (5-10), the Vedernikov/Escoffier and Boyd criterion for physical stability as applicable to prismatic quasi-simple channels. Thus our general Fourier series analysis is capable of producing results identical to those obtained by other methods of linear analysis, but (6-28) is more general than (5-10) in that the effect of shape and width variations and lateral flows on the physical stability criteria are included.

Case 2 $|F|$ negligible, $|G_4|/\sigma$ appreciable, $|\gamma_1|/\sigma$ negligible, i.e. near-stationary water in a non-prismatic channel with negligible lateral inflow. Note that $FG_4/\sigma$ is negligible compared with $G_4/\sigma$.

(6-28) becomes

$$\frac{|G_4|}{\sigma} \leqslant 2 \tag{6-29}$$

Now it is clear from (6-22) that if F is sufficiently small in turbulent flow this criterion must be infringed in any non-prismatic channel. In estuarine reaches of rivers, for instance, the velocity of the flow may tend to zero during the tidal cycle, in which case we would predict instability from (6-29). Similarly we might predict instability from (6-29) in the near stationary waters of a lake. It is important in this case to bear in mind the limitations of our basic theory of Chapter 2, where we restricted our attention to "long" waves, whose amplitude is small compared with their depth and wave-length. Thus (6-29) does not mean that any small displacement of the water surface in a stationary body of water will grow without limit in an unstable manner. Such growth would contradict everyday experience. What (6-29) does mean is that any "long" wave imposed on a near stationary body of water will steepen locally in an unstable manner as it progresses along the channel if the shape and width variation has more influence than the frictional resistance. This is in accord with experience where "long" waves are generated by tides in funnel shaped estuaries such as that of the Severn River in England, or by steady wind influence in lakes, or even by earthquakes in the sea.

In the same way we might regard the particular physical instability which results if (5-10) is infringed as an instability in the long kinematic wave caused by a flood (Henderson (1966) p.369), rather than simply as an instability in uniform flow as has generally been implied. Otherwise, if the flow was exactly uniform, one would expect a small disturbance to behave in the same way to an observer "following the fluid" whatever the Froude Number.

<u>Case 3</u> $|F|$ appreciable, $|G_4|/\sigma$ appreciable, $|\gamma_1|/\sigma$ negligible, i.e. an appreciable flow through a non-prismatic channel with negligible lateral inflow.

(6-28) becomes

$$\left|4mF - 2k\frac{G_4}{\sigma}\right| \leqslant 4k + (1 - 2k)F\frac{G_4}{\sigma} \tag{6-30}$$

$$\text{(a)}\quad 4m|F| \geqslant 2k\frac{|G_4|}{\sigma}$$

Note this means that satisfaction of (5-10) implies satisfaction of (6-29). Now F takes the same sign as V, while

$G_4$ takes the same sign as $H_s$, i.e. positive if the area increases with s. If F and $G_4$ have the same sign, (6-30) can be written

$$\left|mF\right| \leqslant k + \frac{\left|G_4\right|}{4\sigma}(2k-1)\left[\frac{2k}{2k-1} - \left|F\right|\right] \tag{6-31}$$

If F and $G_4$ have opposite signs, (6-30) becomes

$$\left|mF\right| \leqslant k - \frac{\left|G_4\right|}{4\sigma}(2k-1)\left[\frac{2k}{2k-1} - \left|F\right|\right] \tag{6-32}$$

At the point of incipient instability given by (5-10), $|F| = k/m$ approx., so that

$$\frac{2k}{2k-1} - \left|F\right| = \frac{2k}{2k-1} - \frac{k}{m} = \frac{2k\left[m-(k-\frac{1}{2})\right]}{m(2k-1)} \tag{6-33}$$

Thus if $m = k-\frac{1}{2}$ shape and width variations ($G_4$) will have no effect on the physical stability criterion, which will remain (5-10).

(b) $4m\left|F\right| < 2k\frac{\left|G_4\right|}{\sigma}$

Note this means that satisfaction of (6-29) implies satisfaction of (5-10).

(6-30) becomes, if F and $G_4$ have the same signs,

$$\frac{\left|G_4\right|}{\sigma} \leqslant 2 + \frac{2\left|F\right|}{k}\left[m - \frac{\left|G_4\right|}{2\sigma}(k - \tfrac{1}{2})\right] \tag{6-34}$$

If F and $G_4$ have opposite signs (6-30) becomes

$$\frac{\left|G_4\right|}{\sigma} \leqslant 2 - \frac{2\left|F\right|}{k}\left[m - \frac{\left|G_4\right|}{2\sigma}(k - \tfrac{1}{2})\right] \tag{6-35}$$

Thus if $m = k-\frac{1}{2}$ an increase in F will have no effect on the stability criterion (6-29), although it will usually have a stabilizing effect by decreasing $G_4/\sigma$.

In both alternatives 3(a) and 3(b), the criteria (5-10) and (6-29) respectively are unaltered if $m = k-\frac{1}{2}$. This corresponds with $k = 2\frac{1}{2}$ for laminar flow and $k = 1$ to $\frac{7}{6}$ for turbulent flow. If $m > k - \frac{1}{2}$, as we would expect in most cases of laminar flow, increasing width in the same direction as the velocity makes the criteria (5-10) and (6-29) somewhat conservative, while area decreasing in the same direction as the velocity makes (5-10) and (6-29) underconservative. If $m < k - \frac{1}{2}$ the opposite happens.

However, because $|m - (k-\frac{1}{2})|$ is rarely large, (5-10) is adequate for Case 3(a) as long as (6-21) and (6-24) hold. Similarly in Case 3(b), (6-29) is adequate if (6-24) holds, because if $|G_4|/2\sigma$ is of O(1) then $|F|$ must be small.

If (6-21) or (6-24) do not hold, then the appropriate criterion must be selected from (6-31), (6-32), (6-34) and (6-35).

<u>Case 4</u> $|F|$ appreciable, $|G_4|/\sigma$ negligible, $|\gamma_1|/\sigma$ appreciable, i.e. an appreciable flow through a near prismatic channel subject to lateral flow.

(6-28) becomes

$$\left|4(mF + k\frac{\gamma_1 U}{\sigma c})\right| \leqslant 4k(1 + \frac{\gamma_1}{\sigma}) + 2(2k-1)\frac{\gamma_2}{\sigma} \tag{6-36}$$

(a) $m|F| \geqslant k\,\dfrac{\gamma_1 U}{\sigma c}$

If F is positive, (6-36) becomes, writing $\gamma_1 = \gamma_2 = \gamma$

$$|mF| \leqslant k + \frac{\gamma}{2\sigma}\left[2k(2 - \frac{U}{c}) - 1\right] \tag{6-37}$$

If F is negative, (6-36) becomes

$$|mF| \leqslant k + \frac{\gamma}{2\sigma}\left[2k(2 + \frac{U}{c}) - 1\right] \tag{6-38}$$

If we take U = V-u from (2-20), and u = 0, (6-37) and (6-38) can both be expressed

$$|mF| \leqslant k + \frac{\gamma}{2\sigma}\Big[2k(2 - |F|) - 1\Big] \qquad (6\text{-}39)$$

If $|F| = k/m$ approx., u, the s velocity component of the lateral flow, will commonly be small compared with V, so that (6-39) can generally be used in place of (6-37) and (6-38). Now $2k(2 - \frac{k}{m}) - 1 = 0$ when $k = m \pm (m^2 - \frac{1}{2}m)^{\frac{1}{2}}$

As $k \geqslant 1$ the only possibilities of (5-10) being unaffected by lateral flows are therefore $k = 2 + \sqrt{3}$ for laminar flow (m=2), and k = 1 for turbulent flow ($m = \frac{2}{3}$). However, $\frac{1}{2}(2k(2-|F|)-1)$ will rarely be larger than O(1) if $|F| = k/m$ approx., so that, as $\gamma/\sigma$ is usually small (see (6-15) and (6-16)), the moderate lateral inflows of Case 4(a) normally have little influence on the criterion (5-10).

(b) $m|F| < k\left|\frac{\gamma_1 U}{\sigma c}\right|$

We take it that q and hence $\gamma_1, \gamma_2$ are positive in this case as a lateral outflow sufficiently large for $\gamma/\sigma$ to be appreciable is likely to have u = V approx., so that from (2-20) U would be small and Case 4(a) would apply. Further, it seems that any appreciable lateral outflow cannot be formulated as in Chapter 2 (see Section 6.4) so that such outflows cannot be introduced to (6-28). If U and F have the same sign (6-36) becomes

$$\frac{|U|}{c} \leqslant (1 - \frac{m|F|}{k})\frac{\sigma}{\gamma} + 2 - \frac{1}{2k} \qquad (6\text{-}40)$$

If U and F have opposite signs, (6-28) becomes

$$\frac{|U|}{c} \leqslant \left(1 + \frac{m|F|}{k}\right)\frac{\sigma}{\gamma} + 2 - \frac{1}{2k} \qquad (6\text{-}41)$$

From (2-20), -U is the velocity of the inflow as it appears to be an observer "following the fluid" at velocity V and hence a large magnitude of U is more likely if the inflow enters in the opposite direction to the main flow i.e. if u and V take opposite signs. In this case U and V take the same sign and hence (6-40) is more likely to be critical than (6-41).

Alternative 4(b) becomes important only in the unlikely case that $\gamma_1/\sigma$ is large, when a high velocity lateral inflow would presumably build up an unstable long wave profile.

In general, however, $\gamma_1/\sigma$ will be small as may be seen from (6-14) and (6-16), so that the lateral flow terms in (6-28) will rarely be important.

Summary

We have applied criterion (6-28) to a number of special cases, and predicted that a long wave which possesses flow parameters infringing this criterion will steepen locally in an unstable manner until the analysis of Chapter 2 does not hold. We have shown that (6-28) takes the simple form (5-10) if the frictional resistance, which can be associated with the Froude Number, predominates over channel shape and width variations (Case 1), and the simple form (6-29) if the channel shape and width variations are predominant (Case 2). If frictional resistance and shape and width variations are both appreciable

(6-28) can be expressed either as a modified form of (5-10), in Case 3(a), or as a modified form of (6-29) in Case 3(b).

The effect of the lateral flow terms in (6-28) is considered in Case 4, and is shown to be appreciable only in the exceptional case of an intense lateral inflow entering the channel at a considerable velocity relative to the main flow velocity. In other cases, roughly speaking, lateral flow terms are either relatively small or have self-compensating influences on either side of (6-28).

Although the violation of (6-28) implies physical instability it cannot be concluded that satisfaction of (6-28) implies that the physical flow must conform to the assumptions of Chapter 2, as any strong disturbance of the flow, such as the rapid increase in the depth at the end of the channel described by Henderson (1966) Pp.297-304, may cause disturbances in the surface profile. Thus a distinction must be drawn between externally imposed disturbances, which should be traceable through a difference solution by special techniques such as that suggested by Abbott (1966), Section 4.4, and the inherent physical instability discussed in this section.

## 6.4 Interpretation of Stability Criteria for $\theta = 0$

We restate the criteria developed for near-prismatic channels in Section 5.10, using (6-11) and (6-12). Thus (5-70) and (5-71) become, dropping terms in $H_s$,

$$\left| \frac{2k - 2\alpha_2(2k-1)\gamma_2\Delta t}{2k + 2\alpha_1(2k-1)\gamma_2\Delta t} \right| \leqslant 1 \qquad (6\text{-}42)$$

$$\left|\frac{1 - 2\alpha_2(\sigma + \gamma_1)\Delta t}{1 + 2\alpha_1(\sigma + \gamma_1)\Delta t}\right| \leqslant 1 \qquad (6\text{-}43)$$

Now for $\theta = 0$, neither $\alpha_1$ nor $\alpha_2$ can be negative, and $\alpha_1 + \alpha_2 = 1$. Thus (6-42) must be infringed for all $\Delta t$ if q is negative, assuming we restrict $2k + 2\alpha_1(2k-1)\gamma_2\Delta t > 0$. This is necessary as we clearly cannot permit $\lambda$ to tend to infinity. As (6-42) applies for all $\Delta t$, including $\Delta t$ small, it is clear that physical instability in the flow is predicted if q is negative, which confirms the instability for lateral outflows discussed in Section 4.2. This does not appear to conform to everyday experience, although the effect of lateral outflow on long waves has hardly been closely examined. However the instability is probably caused in this case by an improperly posed lateral outflow, because we have assumed that q is specified quite independently of the flow solution. This is reasonable for many lateral inflows, but all lateral outflows must physically depend to some extent on the depth of water above the outflow region as well as the local velocity of flow. Thus if we attempt to model lateral outflow by a separately specified parameter q in a numerical solution we may introduce instability not present in the actual flow, and this seems to account for (6-42).

Assuming $\gamma_1 \geqslant 0$ then, we can satisfy (6-43) for all $\Delta t$ if and only if $\alpha_1 \geqslant \alpha_2$ for $\theta = 0$, since $\alpha_1 + \alpha_2 = 1$ and $\alpha_1, \alpha_2 \geqslant 0$. In other words all difference schemes, including implicit schemes, will become unstable for $\Delta t$ large unless at least as much weight is put on the forward time line $t = t + \Delta t$ as the backward time line $t = t$ in the evaluation of non-homogeneous terms.

Note that criterion (6-43) incorporates the special criteria developed in the second example of Section 4.2. This is not surprising when we remember (see Section 5.9) that the case $\theta = 0$ represents the variations, or errors, as constant along each time line, which was precisely the assumption made in Section 4.2.

### 6.5 A General Explicit Formulation

Substituting from (6-11) and (6-12) into (5-46) to (5-49) we can write

$$\mu_{11} - \mu_{22} = (2kG_2 - G_3)\Delta t(\alpha_2\Psi_1 + \alpha_1\Psi_2) - (G_1 + 2kG_4)\Delta t(\alpha_2\Phi_1 + \alpha_1\Phi_2) \quad (6\text{-}44)$$

$$\mu_{12} = \Psi_1\Phi_2 - \Phi_1\Psi_2 - G_4\Delta t(\alpha_2\Psi_1 + \alpha_1\Psi_2) + G_2\Delta t(\alpha_2\Phi_1 + \alpha_1\Phi_2) \quad (6\text{-}45)$$

$$\mu_{21} = 4k^2(\Psi_1\Phi_2 - \Phi_1\Psi_2) + 2kG_1\Delta t(\alpha_2\Psi_1 + \alpha_1\Psi_2) + 2kG_3\Delta t(\alpha_2\Phi_1 + \alpha_1\Phi_2) \quad (6\text{-}46)$$

As shown by equation (5-44), if we wish to examine the general behaviour of the eigenvalues with variations in $\theta$, we must be able to find the square root of $\Omega$. We simplify the expression for $\sqrt{\Omega}$ by setting $\psi_1 = \psi_3 = 0$ in (5-36) and (5-38), which from (5-26) to (5-29) means that we are now excluding fully implicit formulations. Thus from (5-36) and (5-38)

$$\Psi_1 = x_1 \quad (6\text{-}47)$$

$$\Phi_1 = 0 \quad (6\text{-}48)$$

We also set

$$X = \alpha_2\chi_1 + \alpha_1\Psi_2 \tag{6-49}$$

Thus (6-44), (6-45) and (6-46) simplify to

$$\mathcal{M}_{11} - \mathcal{M}_{22} = (2kG_2 - G_3)X\Delta t - (G_1 + 2kG_4)\alpha_1\Phi_2\Delta t \tag{6-50}$$

$$\mathcal{M}_{12} = -G_4X\Delta t + (\chi_1 + \alpha_1G_2\Delta t)\Phi_2 \tag{6-51}$$

$$\mathcal{M}_{21} = 2kG_1X\Delta t + 2k(2k\chi_1 + \alpha_1G_3\Delta t)\Phi_2 \tag{6-52}$$

The expression for $\Omega$ is simplified still more if we neglect either $G_1$ or $G_4$. We shall neglect $G_4$ in this analysis as we are primarily interested in flows of all Froude Numbers in near prismatic channels, but the proof follows similar lines if the Froude number (and lateral inflow) is taken as small and $G_1$ is dropped. The same applies to Section 5.10, where (5-70) and (5-71) are equally obtainable by dropping $\tau_1$ and $\tau_2$ and retaining $\nu_3$ and $\nu_4$, as by dropping $\nu_3$ and $\nu_4$ and retaining $\tau_1$ and $\tau_2$. Note that we are not excluding limited shape and width variations, as if F is not small such variations must be considerable before $G_4$ becomes important, as shown by (6-21) and (6-24).

The equations (5-46) to (5-50) now become

$$\mathcal{M}_{11} = 2k\chi_1\Psi_2 + 2kG_2\alpha_1\Psi_2\Delta t - G_3\alpha_2\chi_1\Delta t - G_2G_3\alpha_1\alpha_2\Delta t^2 \tag{6-53}$$

$$\mathcal{M}_{22} = 2k\chi_1\Psi_2 + G_3\alpha_1\Psi_2\Delta t - 2kG_2\alpha_2\chi_1\Delta t + G_1\alpha_1\Phi_2\Delta t - G_2G_3\alpha_1\alpha_2\Delta t^2 \tag{6-54}$$

$$\mathcal{M}_{12} = (\chi_1 + \alpha_1G_2\Delta t)\Phi_2 \tag{6-55}$$

$$\mathcal{M}_{21} = 2kG_1X\Delta t + 2k(2k\chi_1 + \alpha_1G_3\Delta t)\Phi_2 \tag{6-56}$$

$$\det(M_1) = (2k\varkappa_1 + \alpha_1 G_3 \Delta t)(\varkappa_1 + \alpha_1 G_2 \Delta t) \tag{6-57}$$

From (5-45)

$$\Omega = \tfrac{1}{4}(2kG_2 - G_3)^2 X^2 \Delta t^2 + G_1(2k\varkappa_1 + \tfrac{1}{2}\alpha_1(2kG_2 + G_3)\Delta t)X\Delta t\Phi_2 + (\tfrac{1}{4}\alpha_1{}^2 G_1{}^2 \Delta t^2 + 2k(2k\varkappa_1 + \alpha_1 G_3 \Delta t)(\varkappa_1 + \alpha_1 G_2 \Delta t))\Phi_2{}^2 \tag{6-58}$$

## 6.6 Initial Restrictions

Equation (5-42) suggests that $\lambda$ tends to infinity as $\det(M_1)$ tends to zero, but we have not quite established this as it is just possible that $\Lambda$ tends to zero with $\det(M_1)$ in such a way that $\lambda$ still satisfies the stability condition (4-56). We therefore now investigate the behaviour of $\Lambda$ as $\det(M_1)$ tends to zero in order to show that $\lambda$ must violate (4-56) in this case. This leads to initial restrictions on $\varkappa_1$, $\alpha_1$ and $\Delta t$.

Let

$$L_1 = \tfrac{1}{2}(2kG_2 - G_3) = 2k(\sigma + \gamma_1 - \gamma_2) + \gamma_2 \tag{6-59}$$

$$L_2 = \tfrac{1}{2}G_1 = 2(mF\sigma + k\gamma_1 \tfrac{U}{1c}) \tag{6-60}$$

$$\Gamma_1 = 2k\varkappa_1 + \tfrac{1}{2}\alpha_1(2kG_2 + G_3)\Delta t \tag{6-61}$$

$$\Gamma_2^2 = \tfrac{1}{4}\alpha_1{}^2 G_1{}^2 \Delta t^2 + 2k \det(M_1) \tag{6-62}$$

so that

$$\Omega = L_1{}^2 \Delta t^2 X^2 + 2L_2 \Delta t \Gamma_1 X \Phi_2 + \Gamma_2^2 \Phi_2{}^2 \tag{6-63}$$

Now

$$\tfrac{1}{2}(\mu_{11}+\mu_{22}) = (2k\chi_1 + \alpha_1 G_3\Delta t)(\Psi_2 - \alpha_2 G_2\Delta t) + L_1\Delta tX + \alpha_1 L_2\Delta t\Phi_2 \quad (6\text{-}64)$$

$$= (\chi_1 + \alpha_1 G_2\Delta t)(2k\Psi_2 - \alpha_2 G_3\Delta t) - L_1\Delta tX + \alpha_1 L_2\Delta t\Phi_2 \quad (6\text{-}65)$$

From (6-62), as $\det(M_1) \longrightarrow 0$, $\Gamma_2^{\,2} \longrightarrow \alpha_1^{\,2}L_2^{\,2}\Delta t^2$ (6-66)

From (6-57), $\det(M_1) \longrightarrow 0$ if either of its factors tends to zero.

We first let

$$2k\chi_1 + \alpha_1 G_3\Delta t \longrightarrow 0 \quad (6\text{-}67)$$

Hence from (6-61) $\Gamma_1 \longrightarrow \alpha_1 L_1\Delta t$

Thus from (6-63) $\Omega \longrightarrow (L_1\Delta tX + \alpha_1 L_2\Delta t\Phi_2)^2$

Therefore from (5-44), (6-64)

$$\Lambda_{1,2} \longrightarrow (L_1\Delta tX + \alpha_1 L_2\Delta t\Phi_2)(1 \pm 1) \quad (6\text{-}68)$$

Next we let $\chi_1 + \alpha_1 G_2\Delta t \longrightarrow 0$ (6-69)

From (6-61) $\Gamma_1 \longrightarrow -\alpha_1 L_1\Delta t$

Thus from (6-63) $\Omega \longrightarrow (-L_1\Delta tX + \alpha_1 L_2\Delta t\Phi_2)^2$

Therefore from (5-44), (6-65)

$$\Lambda_{1,2} = (-L_1\Delta tX + \alpha_1 L_2\Delta t\Phi_2)(1 \pm 1) \quad (6\text{-}70)$$

Thus from (6-68), (6-70) there is in general one non-zero root of $\Lambda$ when $\det(M_1) \longrightarrow 0$ and hence from (5-42), (4-56) we have our initial restrictions

$$2k\chi_1 + \alpha_1 G_3\Delta t > 0 \quad (6\text{-}71)$$

$$\chi_1 + \alpha_1 G_2\Delta t > 0 \quad (6\text{-}72)$$

Because $\sigma$ and $\gamma$ are positive, $G_2$ and $G_3$ are positive, so the main effect of (6-71) and (6-72) is virtually to restrict $\chi_1$ to the value 1, as any normal averaging process such as that represented by $\frac{1}{2}(1 + \cos\theta)$ causes $\chi_1$ to approach zero for some value of $\theta$. Similarly (6-71) and (6-72) suggest that $\alpha_1$ should be independent of $\theta$, as limits on $\Delta t$ are introduced if $\alpha_1$ can take negative values.

## 6.7 Evaluation of the Square Root

We now investigate $\sqrt{\Omega}$ for all $\Delta t$ and $\theta$. This is necessary for a complete stability analysis, as may be seen from (5-44).

We can relate $\Gamma_1$ and $L_1$ to the factors of $\det(M_1)$ by

$$\Gamma_1 - \alpha_1 L_1 \Delta t = 2k\chi_1 + \alpha_1 G_3 \Delta t \tag{6-73}$$

$$\Gamma_1 + \alpha_1 L_1 \Delta t = 2k(\chi_1 + \alpha_1 G_2 \Delta t) \tag{6-74}$$

Thus, using (6-57), (6-62) and (6-60)

$$\Gamma_1^2 - \Gamma_2^2 = \alpha_1^2 (L_1^2 - L_2^2)\Delta t^2 \tag{6-75}$$

Therefore from (6-63) $\Omega$ is a perfect square when $L_1 = L_2$ and it appears that $\sqrt{\Omega}$ has different properties when $L_1 > L_2$ and when $L_1 < L_2$. Now $L_1$ and $L_2$ incorporate the $L_1$, $L_2$ defined in (4-37) for the special case of Chapter 4, except for a factor of 2. Just as in Section 4.7 it will be necessary (see Appendix B) to our stability analysis to assume

$$|L_2| \leqslant L_1 \tag{6-76}$$

that is, using (6-59), (6-60) and multiplying throughout by $2/\sigma$

$$\left|4(mF + k\frac{\gamma_1 U}{\sigma c})\right| \leqslant 4k(1 + \frac{\gamma_1}{\sigma}) - 2(2k-1)\frac{\gamma_2}{\sigma}$$

This is identical with the physical stability criterion (6-36) except for the sign of the term in $\gamma_2/\sigma$. Now, as shown in Section 6.4, if $\gamma_2/\sigma$ is appreciable $\gamma_2$ must be positive for stability, so that (6-76) is more conservative than (6-36), but as indicated in Section 6.2, $\gamma_2/\sigma$ is important only if the lateral inflow along a short length of channel is comparable with the flow in the channel itself.

Thus in all but exceptional cases (6-76) can be regarded as necessary for physical flow stability.

As is proved in Appendix B, we can use (6-76) in (6-75), and our initial restrictions (6-71) and (6-72) to give

$$\sqrt{\Omega} = \kappa_1 L_1 \Delta t X + \kappa_2 \Gamma_2 \Phi_2 \tag{6-77}$$

where $\kappa_1$ and $\kappa_2$ are real variables such that

$$\left|\kappa_{1,2}\right| \leqslant 1 \tag{6-78}$$

## 6.8 Evaluation of the Eigenvalues

From (5-44), (6-77)

$$\begin{aligned}\Lambda &= \tfrac{1}{2}(\mu_{11} + \mu_{22}) + \kappa_1 L_1 \Delta t X + \kappa_2 \Gamma_2 \Phi_2 \\ &= \left[2k\chi_1 + (L_1(1+\kappa_1)+G_3)\alpha_1 \Delta t\right]\Psi_2 + \left[\kappa_2 \Gamma_2 + L_2 \alpha_1 \Delta t\right]\Phi_2 \\ &\qquad - \alpha_2 \chi_1 \Delta t\left[L_1(1-\kappa_1) + G_3\right] - \alpha_1 \alpha_2 G_2 G_3 \Delta t^2\end{aligned}$$

using (6-53), (6-54) and (6-49). Thus using (5-37), (5-39) and (6-73)

$$\Lambda = 2k\chi_1\chi_2 + \Delta t\left[(\alpha_1\chi_2-\alpha_2\chi_1)(L_1+G_3) + (\alpha_1\chi_2+\alpha_2\chi_1)K_1L_1 - \alpha_1\alpha_2G_2G_3\Delta t\right]$$

$$- \tfrac{1}{2}(\psi_2-\psi_4)\left[c(\Gamma_1+K_1\alpha_1L_1\Delta t) + V(K_2\Gamma_2+\alpha_1L_2\Delta t)\right]$$

$$- \tfrac{1}{2}(\psi_2+\psi_4)\left[V(\Gamma_1+K_1\alpha_1L_1\Delta t) + c(K_2\Gamma_2+\alpha_1L_2\Delta t)\right] \qquad (6\text{-}79)$$

$$= (2k\chi_1 + \alpha_1G_3\Delta t)(\chi_1 + \alpha_1G_2\Delta t)\lambda \qquad (6\text{-}80)$$

using (5-42) and (6-57). We have arranged (6-79) as such because in most formulations only $\psi_2+\psi_4$ contains an imaginary part.

By investigating the maxima of $\lambda$ as given by (6-79) and (6-80) we can use the stability condition (4-56) to establish the stability properties for all $\Delta t$, $\Delta s$ of all difference schemes satisfying our assumptions, i.e. non-implicit, regular net, one cycle, two level difference schemes. The maxima of $\lambda$ clearly depend on the relationship with $\theta$ of $\chi_1$, $\chi_2$, $\alpha_1$, $\alpha_2$, $K_1$, $K_2$, $\psi_2$, $\psi_4$, $\Gamma_1$ and $\Gamma_2$.

Any centred space average at time $(b+1)\Delta t$ for $\chi_1$ and $\alpha_1$ introduces a term in $\cos\theta$, so that $\chi_1$ and $\alpha_1$ then reach a maximum when $\theta = 0$. However, these maximum values are retained for all $\theta$ if $\chi_1$ and $\alpha_1$ result from terms evaluated only at $(a\Delta s, (b+1)\Delta t)$, as they are then independent of $\theta$. As may be deduced by comparing (5-52) and (5-53) with (6-79) and (6-80), the larger the value of $\chi_1$ for given $\theta$, the smaller the magnitude of $\lambda$. Similarly the larger the value of $\alpha_1$ the smaller the magnitude of $\lambda$, because $G_2$ and $G_3$ are both restricted to positive values. Thus making $\chi_1$ and $\alpha_1$ independent of $\theta$ appears to have a stabilizing effect and hence we shall discuss only difference schemes which use constant values of $\chi_1$ and $\alpha_1$. We need therefore only investigate maxima of $|\Lambda|$ with respect to variation in $\theta$ as these coincide with maxima in $|\lambda|$.

Maxima of $|\Lambda|$ occur when $\theta$ and $K_{1,2}$ take certain critical values. Such values of $K_{1,2}$ are functions of the flow properties, $\Delta t$ and the critical value of $\theta$, but a conservative estimate of each of these values is in all cases the more critical of +1 or -1, from (6-78). Thus we can make the conservative assumption that $|K_{1,2}| = 1$ for all $\theta$ which enables us to treat $K_{1,2}$ as independent of $\theta$, and this greatly simplifies the relationship between $|\Lambda|$ and $\theta$. Note that this assumption becomes exact as $\Delta t \longrightarrow 0$ because from (6-63) $\Omega \longrightarrow \Gamma_2^2 \phi_2^2$ and hence from (6-77) $|K_2| \longrightarrow 1$ and $K_1$ is no longer important.

In investigating the maxima of $|\Lambda|$ from (6-79) it should be remembered that only $\chi_1$, $\chi_2$, $\psi_2$, $\psi_4$, $\alpha_1$, $\alpha_2$ and $\Gamma_1$, $\Gamma_2$ vary with the difference formulation, because k, $G_1$, $G_2$, $G_3$ and $L_1$, $L_2$ are properties of the solution.

It is more convenient to treat individual difference schemes from this stage than to carry the general analysis further, because any simplifications are generally related to the character of the individual schemes. However, the analyses follow similar lines as we illustrate with examples.

## 6.9 General Stability Analysis of the Staggered Scheme

We shall compare the stability properties obtained for the Lax staggered scheme with those derived for the same scheme in Chapter 4.

$\chi_1 = 1$, $\chi_2 = \cos\theta$, $\psi_2 = \psi_4 = i\frac{\Delta t}{\Delta s}\sin\theta$, $\alpha_1 = 0$, $\alpha_2 = \cos\theta$.

Thus $\Gamma_1 = \Gamma_2 = 2k\chi_1 = 2k$ from (6-61), (6-62) and (6-57).

Therefore, from (6-79)

$$\Lambda = 2k\cos\theta + \Delta t\left[-\cos\theta\,(L_1+G_3) + \cos\theta\,K_1L_1\right] - i\frac{\Delta t}{\Delta s}\sin\theta\left[2kV+2kK_2c\right]$$

$$= \cos\theta\left[2k + \Delta t(K_1L_1 - L_1 - G_3)\right] - i2k\sin\theta\left[W + K_2Z\right] \qquad (6\text{-}81)$$

introducing W and Z as defined in (A-18).

$$|\Lambda|^2 = \cos^2\theta\left[2k - (L_1+G_3-K_1L_1)\Delta t\right]^2 + 4k^2\sin^2\theta\left[W + K_2Z\right]^2 \qquad (6\text{-}82)$$

Hence, treating $K_{1,2}$ as independent of $\theta$ as discussed in Section 6.8

$$\frac{d|\Lambda|^2}{d\theta} = -2\cos\theta\sin\theta\left[2k - (L_1+G_3-K_1L_1)\Delta t\right]^2 + 2\sin\theta\cos\theta\,4k^2\left[W + K_2Z\right]^2$$

$$\frac{d^2|\Lambda|^2}{d\theta^2} = 2(\cos^2\theta - \sin^2\theta)\left[4k^2(W + K_2Z)^2 - (2k - (L_1+G_3-K_1L_1)\Delta t)^2\right]$$

It can be seen that the turning points occur in the way described in Chapter 4, and we then investigate $|\lambda|^2$ at such points. In the case $\cos\theta = 0$

$$|\lambda|^2 = \frac{|\Lambda|^2}{(\det(M_1))^2} = \frac{4k^2\sin^2\theta\,(W + K_2Z)^2}{4k^2}$$

$$= (W + K_2Z)^2$$

Because $|K_2| \leqslant 1$ we know that

$$|\lambda|^2 \leqslant (W \pm Z)^2 \qquad (6\text{-}83)$$

In the case $\sin\theta = 0$

$$|\lambda|^2 = \frac{|\Delta|^2}{(\det(M_1))^2} = \frac{\cos^2\theta\,(2k - (L_1 + G_3 - \kappa_1 L_1)\Delta t)^2}{4k^2}$$

$$\leqslant \max\left[(1 - \frac{G_3\Delta t}{2k})^2,\ (1 - (2L_1 + G_3)\frac{\Delta t}{2k})^2\right]$$

$$\leqslant \max\left[(1 - \frac{2k-1}{k}\gamma_2\Delta t)^2,\ (1 - 2(\sigma+\gamma_1)\Delta t)^2\right] \qquad (6\text{-}84)$$

using (6-9), (6-59) and (6-8).

The other possible turning point occurs when

$$\left[2k - (L_1 + G_3 - \kappa_1 L_1)\Delta t\right]^2 = 4k^2(W + \kappa_2 Z)^2$$

and it can be seen from (6-82) that this case is covered by (6-83) and (6-84). Thus, combining (6-83) and (6-84) we can say for all $\theta$

$$|\lambda|^2 \leqslant \max\left[(|W| + Z)^2,\ (1 - \frac{2k-1}{k}\gamma_2\Delta t)^2,\ (1 - 2(\sigma+\gamma_1)\Delta t)^2\right] \qquad (6\text{-}85)$$

which reduces to (4-54) in the special case W +ve, k = 1, $\gamma_1 = \gamma_2 = \gamma$ taken in Chapter 4.

The stability criterion (4-56) gives three conditions, from (6-85), which must be satisfied independently. These are

$$(|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \qquad (6\text{-}86)$$

$$\gamma_2 \geqslant 0 \qquad (6\text{-}87)$$

$$(\sigma + \gamma_1)\Delta t \leqslant 1 \qquad (6\text{-}88)$$

Note that (6-87) is sufficient to ensure that $\sigma+\gamma_1$ is not negative as $\eta$ is positive.

Because there is no term in $\cos\theta$ (as distinct from $\cos^2\theta$)

in (6-82) we could take $\sin\theta = 0$ to mean $\theta = 0$, so that (6-87) and (6-88) follow at once from (6-42) and (6-43) respectively.

Our final stability conditions (6-86), (6-87) and (6-88) for the staggered scheme clearly incorporate the corresponding (4-58), (4-59) and (4-60), and were readily obtained from (6-79) and (6-80). This illustrates the value of (6-79) and (6-80) as a general expression of the eigenvalues required for use with the stability condition (4-56).

## 6.10 Types of Numerical Instability

In the case $\sin\theta = 0$ we can attach a sign to the magnitude $|\lambda|$ because the imaginary part of $\lambda$ disappears. From (6-81), when $\sin\theta = 0$,

$$\lambda = \cos\theta\left[1 - \frac{\Delta t}{2k}(L_1 - \kappa_1 L_1 + G_3)\right] \tag{6-89}$$

Now it is reasonable to associate monotonic instability with $\lambda > 1$ and oscillatory instability with $\lambda < -1$ by analogy with the influence of the simple amplification factors discussed in Section 4.2, but the adjectives "monotonic" and "oscillatory" here apply only to the "timewise" variation from the true solution, that is, the variation with time of a solution along a line s = constant on the s-t plane. In Section 4.2 the "spacewise" variation from the true solution along a line t = constant was taken as constant, which as explained in Section 5.9 is the equivalent of taking $\theta = 0$ in our Fourier Series analysis, but if we take $\theta = \pm\pi$ we represent a variation which is "spacewise" oscillatory, taking alternating positive and negative values of constant magnitude at successive mesh points along a line t = constant.

Thus if the variations from the true solution are initially constant ($\cos\theta = 1$) along a line t = const. in a staggered scheme solution*, from (6-89) we might expect them to grow monotonically with time if (6-87) is violated and in an oscillatory manner with time if (6-88) is violated. Alternatively if the initial variations are constant in magnitude but alternate in sign ($\cos\theta = -1$) we might expect oscillatory growth with time of each variation if (6-87) is violated and monotonic growth of each variation if (6-88) is violated.

This illustrates four distinct idealised types of instability, which might be called respectively monotonic-monotonic, oscillatory-monotonic, oscillatory-oscillatory and monotonic-oscillatory instability. The first two pure types are of some practical interest as almost constant errors along a line t = const. will commonly arise from the discretization of a continuous solution into a difference solution, as in the first example of Section 4.2; the last two types are likely to occur only in combination with the first two.

The nature of these four types of instability indicates

---

* Although the staggered scheme provides solutions only at alternate points of a rectangular mesh, for the sake of this illustration we assume that the solution at all points of the rectangular mesh is desired and that the staggered scheme is therefore applied twice, first to obtain the solution at half the mesh points, then independently to obtain the solution at the remaining mesh points.

the danger of the empirical approach to stability investigation. Such an approach relies heavily on oscillations in the numerical solution to warn of instability, whereas three of the four pure types of unstable solution may not exhibit oscillations if the solution is examined along lines parallel to one axis of the s-t plane, and pure monotonic-monotonic instability does not create oscillations at all.

In practice therefore, because the idealised pure types of instability will rarely occur in isolation, we shall use "oscillatory instability" to mean instability causing oscillations which obviously make the solution unsatisfactory, and "monotonic instability" to mean all other cases of instability.

### 6.11 General Stability Analysis of the "Unstable" Scheme

This scheme is described by Liggett and Woolhiser (1967) p.53.

$$\chi_1 = 1,\ \chi_2 = 1,\ \psi_2 = \psi_4 = i\frac{\Delta t}{\Delta s}\sin\theta,\ \alpha_1 = 0,\ \alpha_2 = 1.$$

Hence again $\Gamma_1 = \Gamma_2 = 2k$. Therefore, from (6-79)

$$\Lambda = 2k - \Delta t(L_1 + G_3 - K_1L_1) - i2k\sin\theta\,(W + K_2Z)$$

$$|\Lambda|^2 = \left[2k - (L_1 + G_3 - K_1L_1)\Delta t\right]^2 + 4k^2\sin^2\theta\,(W + K_2Z)^2$$

The maximum clearly occurs when $\cos\theta = 0$. Again $\det(M_1) = 2k$ so from (6-80)

$$|\lambda|^2 \leqslant \max\left[(1 - \frac{2k-1}{k}\gamma_2\Delta t)^2,\ (1 - 2(\sigma + \gamma_1)\Delta t)^2\right] + (|W| + Z)^2 \qquad (6\text{-}90)$$

It is clear that this scheme cannot satisfy the stability condition (4-56) as $\Delta t$ tends to zero and therefore cannot be recommended, but (6-90) shows that stability is possible for finite $\Delta t$ only if there is lateral inflow, and this is confirmed by the experience of Liggett and Woolhiser with the rising limb of a hydrograph as mentioned in Section 4.1.

This also explains why Morgali and Linsley (1965) obtained apparently stable results when they used the "Unstable" scheme to solve the rising limb of a hydrograph, as $\gamma_2 = q/2A$ would be appreciable in the early part of the hydrograph whereas W + Z would be small. The overshoot above the equilibrium flow which appeared in some of their solutions suggests monotonic instability, which is consistent with (6-90) as $\gamma_2$ would decrease with increase in flow cross-section, whereas W + Z would increase.

## 6.12 General Stability Analysis of a Rectangular Scheme

We now analyse the stability of a rectangular scheme which conforms in all respects to the CIR rectangular scheme discussed in Appendix A, except that we shall not yet restrict $\alpha_1$ and $\alpha_2$ except to assume they are both independent of $\theta$. Thus we incorporate the case $\alpha_1 = 0$, $\alpha_2 = 1$, which is the CIR rectangular scheme, and a semi-explicit case $\alpha_1 = 1$, $\alpha_2 = 0$ in which the unknown solution is used in the formulation of the non-homogeneous terms.

In this scheme $\chi_1 = \chi_2 = 1$ and $\psi_2$ and $\psi_4$ are formulated in order to reflect the domain of dependence of the solution. Thus if $|F| \leqslant 1$,

$$\psi_2 = (1 - e^{-i\theta})\frac{\Delta t}{\Delta s} = (1 - \cos\theta + i \sin\theta)\frac{\Delta t}{\Delta s} \qquad (6\text{-}91)$$

$$\psi_4 = (e^{i\theta} - 1)\frac{\Delta t}{\Delta s} = (\cos\theta + i\sin\theta - 1)\frac{\Delta t}{\Delta s} \qquad (6\text{-}92)$$

If $F > 1$, $\psi_4 = \psi_2$ as defined in (6-91) and if $-F > 1$, $\psi_2 = \psi_4$ as defined in (6-92). From (6-79)

$$\Lambda = P_1 - \tfrac{1}{2}(\psi_2 - \psi_4)\frac{\Delta s}{\Delta t} P_2 - \tfrac{1}{2}(\psi_2 + \psi_4)\frac{\Delta s}{\Delta t} P_3 \qquad (6\text{-}93)$$

where

$$P_1 = 2k + \Delta t\left[(\alpha_1 - \alpha_2)(L_1 + G_3) + (\alpha_1 + \alpha_2)K_1 L_1 - \alpha_1\alpha_2 G_2 G_3 \Delta t\right] \qquad (6\text{-}94)$$

$$P_2 = Z(\Gamma_1 + K_1\alpha_1 L_1\Delta t) + W(K_2\Gamma_2 + \alpha_1 L_2\Delta t) \qquad (6\text{-}95)$$

$$P_3 = W(\Gamma_1 + K_1\alpha_1 L_1\Delta t) + Z(K_2\Gamma_2 + \alpha_1 L_2\Delta t) \qquad (6\text{-}96)$$

and we regard $P_1$, $P_2$, $P_3$ as independent of $\theta$ because we have limits on $K_1$, $K_2$ which are independent of $\theta$.

If $|F| \leqslant 1$,

$$\Lambda = P_1 - (1 - \cos\theta)P_2 - i\sin\theta\, P_3 \qquad (6\text{-}97)$$

If $F > 1$

$$\Lambda = P_1 - (1 - \cos\theta)P_3 - i\sin\theta\, P_3 \qquad (6\text{-}98)$$

If $-F > 1$

$$\Lambda = P_1 + (1 - \cos\theta)P_3 - i\sin\theta\, P_3 \qquad (6\text{-}99)$$

We take first the case $|F| \leqslant 1$, so that

$$|\Lambda|^2 = (P_1 - P_2)^2 + 2(P_1 - P_2)P_2\cos\theta + P_2^2\cos^2\theta + P_3^2\sin^2\theta \qquad (6\text{-}100)$$

$$\frac{d|\Lambda|^2}{d\theta} = 2\sin\theta\left[(P_2 - P_1)P_2 + \cos\theta\,(P_3^2 - P_2^2)\right] \qquad (6\text{-}101)$$

$$\frac{d^2|\Lambda|^2}{d\theta^2} = 2(P_2 - P_1)P_2\cos\theta + 2(\cos^2\theta - \sin^2\theta)(P_3^2 - P_2^2) \qquad (6\text{-}102)$$

Now (6-101) indicates turning points in two possible cases:

Case (i) $\sin\theta = 0$

Case (ii) $\cos\theta = \dfrac{(P_1 - P_2)P_2}{P_3^{\,2} - P_2^{\,2}}$

<u>Case (i)</u> $$\frac{d^2|\Lambda|^2}{d\theta^2} = 2(P_3^{\,2} - P_2^{\,2}) \pm 2(P_2 - P_1)P_2 \tag{6-103}$$

Thus there are: Two maxima if $P_2^{\,2} - P_3^{\,2} > |(P_2 - P_1)P_2|$

A maximum and a minimum if $|P_2^{\,2} - P_3^{\,2}| < |(P_2 - P_1)P_2|$

Two minima if $P_3^{\,2} - P_2^{\,2} > |(P_2 - P_1)P_2|$

<u>Case (ii)</u> $2(P_2 - P_1)P_2 \cos\theta = -2\cos^2\theta\,(P_3^{\,2} - P_2^{\,2})$

so that

$$\frac{d^2|\Lambda|^2}{d\theta^2} = 2\sin^2\theta\,(P_2^{\,2} - P_3^{\,2}) \tag{6-104}$$

As may be seen from (6-101) this case does not arise if

$$|P_2^{\,2} - P_3^{\,2}| < |(P_2 - P_1)P_2|$$

Thus there are: Two minima if $P_2^{\,2} - P_3^{\,2} > |(P_2 - P_1)P_2|$

Two maxima if $P_3^{\,2} - P_2^{\,2} > |(P_2 - P_1)P_2|$

If $P_2^{\,2} - P_3^{\,2} \geqslant 0$ it is clear from (6-104) that Case (ii) represents minima in $|\Lambda|^2$ and can be disregarded as irrelevant to the bounds of the eigenvalues. From (6-95) and (6-96)

$$P_2 = P_4Z + P_5W$$

$$P_3 = P_4W + P_5Z$$

where $P_4 = \Gamma_1 + K_1\alpha_1L_1\Delta t$

$P_5 = K_2\Gamma_2 + \alpha_1L_2\Delta t$

$$P_2^{\ 2} - P_3^{\ 2} = (P_4^{\ 2} - P_5^{\ 2})(Z^2 - W^2) = (P_4^{\ 2} - P_5^{\ 2})Z^2(1 - F^2)$$

Thus we can disregard Case (ii) if $P_4^{\ 2} - P_5^{\ 2} \geqslant 0$, that is, using (6-75), if

$$(1-K_2^{\ 2})\Gamma_2^{\ 2} + 2\alpha_1\Delta t(K_1\Gamma_1 L_1 - K_2\Gamma_2 L_2) + \alpha_1^{\ 2}\Delta t^2(L_1^{\ 2}(1+K_1^{\ 2}) - 2L_2^{\ 2}) \geqslant 0$$

This is clearly satisfied if $\alpha_1\Delta t$ is small, in particular if $\alpha_1 = 0$ as in the CIR rectangular scheme, and intuitively appears reasonable for all $\alpha_1\Delta t$ as $L_1 \geqslant |L_2|$ from (6-76) and $\Gamma_1 \geqslant |\Gamma_2|$ from (6-75). This condition is also satisfied for all $\alpha_1\Delta t$ if $K_1 = 1$, and we now show for the formulation $\alpha_1 = 1$, $\alpha_2 = 0$ that the critical values of $|\Lambda|^2$ in Case (i) occur when $K_1 = 1$, so that they cannot be exceeded by possible maxima from Case (ii).

In Case (i) $\sin\theta = 0$, so $\Lambda$ is pure real and we need not study $|\Lambda|^2$. From (6-97)

$$\Lambda = P_1 \text{ or } P_1 - 2P_2$$

for turning points as $\cos\theta = 1$ or $-1$. Thus when $\alpha_1 = 1$ from (6-94)

$$\Lambda = P_1 = 2k + \Delta t(L_1 + K_1 L_1 + G_3) \tag{6-105}$$

It is clear that because $G_3$ is positive, $\Lambda$ in this expression is always positive for all $\Delta t$, and therefore the critical magnitude of $\Lambda$ occurs when $K_1$ reaches its maximum value of 1. Alternatively, from (6-94), (6-95)

$$\begin{aligned}\Lambda &= P_1 - 2P_2 \\ &= 2k - 2Z\Gamma_1 - 2WK_2\Gamma_2 - \Delta t\left[(K_1(2Z-1) - 1)L_1 + 2WL_2 - G_3\right]\end{aligned} \tag{6-106}$$

Now $\Lambda$ in this expression cannot exceed the maximum positive $\Lambda$ from (6-105) so this $\Lambda$ can only become critical when it is negative. Thus provided 2Z-1 is positive, the critical value of $\Lambda$ from (6-106) occurs when $K_1$ reaches its maximum value of 1. We shall show in a moment that 2Z-1 can safely be assumed positive.

We have therefore proved that Case (ii) turning points can be disregarded for $|F| \leqslant 1$ when $\alpha_1 = 0$ or $\alpha_1 = 1$.

Comparison of (6-98) and (6-99) with (6-97) show that for $|F| > 1$ we can repeat the analysis for $|F| \leqslant 1$ with $\pm P_3$ replacing $P_2$, and therefore from (6-101) Case (ii) does not arise and again we need only examine the real part of $\Lambda$ when $\cos\theta = \pm 1$.

6.13 Stability Conditions for the CIR Rectangular Scheme

In this scheme $\alpha_2' = 1$, $\alpha_1' = 0$ and we have shown that maxima in the eigenvalues in this case occur only when $\sin\theta = 0$ and $\Lambda$ is pure real.

When $\cos\theta = +1$, (6-97), (6-98) and (6-99) all give $\Lambda = P_1$. This is not surprising, because when $\cos\theta = +1$, (6-42) and (6-43) must hold for this scheme, and both are independent of F. $P_1$ has its maximum and minimum values when $K_1 = +1$ and $-1$, from (6-94). When $K_1 = +1$

$$\lambda = \frac{2k - G_3\Delta t}{2k} \qquad (6\text{-}107)$$

from (6-94), (6-80). When $K_1 = -1$, using (6-94), (6-80), and (6-59)

$$\lambda = 1 - G_2\Delta t \qquad (6\text{-}108)$$

Note that we could also derive these extremes in $\lambda$ by substituting $\alpha_1 = 0$, $\alpha_2 = 1$ in (6-42) and (6-43) respectively.

When $\cos\theta = -1$, (6-97) gives

$$\Lambda = 2k + \Delta t(\kappa_1 L_1 - L_1 - G_3) - 4kZ(1 + \kappa_2 F) \qquad (6\text{-}109)$$

(6-98) and (6-99) give respectively

$$\Lambda = 2k + \Delta t(\kappa_1 L_1 - L_1 - G_3) - 4kZ(F + \kappa_2) \qquad (6\text{-}110)$$

$$\Lambda = 2k + \Delta t(\kappa_1 L_1 - L_1 - G_3) - 4kZ(-F - \kappa_2) \qquad (6\text{-}111)$$

Note that $|F| \leqslant 1$ in (6-109), $F > 1$ in (6-110) and $-F > 1$ in (6-111).

Now $\kappa_1 L_1 - L_1 - G_3 \leqslant 0$ for all $\kappa_1$ and $-(1 + \kappa_2 F)$, $-(F + \kappa_2)$, and $-(-F - \kappa_2)$ are negative for all $\kappa_2$ in (6-109), (6-110) and (6-111) respectively, so that $\Lambda \leqslant 2k$, i.e.

$$\lambda \leqslant 1$$

for all $\kappa_1$, $\kappa_2$ when $\cos\theta = -1$. Thus this case is critical only when $\Lambda$ is negative, so we seek $\kappa_1$ and $\kappa_2$ such that $\Lambda$ takes its largest negative magnitude. In all three cases this occurs for $\kappa_1 = -1$, and (6-109), (6-110) and (6-111) all become

$$\Lambda = 2k - 2kG_2\Delta t - 4kZ(1 + |F|)$$

Thus, from (6-80)

$$\lambda = 1 - G_2\Delta t - 2Z(1 + |F|) \qquad (6\text{-}112)$$

so our sufficient stability condition (4-56) becomes, for the CIR rectangular scheme, using (6-107), (6-108) and (6-112)

$$1 - G_2\Delta t - 2Z(1 + |F|) \geqslant -1$$

That is, from (6-8)

$$(\sigma + \gamma_1)\Delta t + (|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \qquad (6\text{-}113)$$

We are assuming that $G_3 \geqslant 0$ because $q \geqslant 0$ as is shown to be required in Section 6.4.

The restriction (6-113) is quite inconvenient, as if $(|V| + c)\Delta t/\Delta s$ tends to 1, $\Delta t$ must tend to zero to satisfy the stability condition, and this explains why Stoker (1957), Pp.489-498, experienced the difficulties with the CIR rectangular scheme which forced him to turn to the staggered scheme.

## 6.14 Stability Conditions for a Semi-Explicit Scheme

We now examine the rectangular scheme in which $\alpha_1 = 1$, $\alpha_2 = 0$, as we have been able to establish that the critical maxima in the eigenvalues in this case occur when $\sin\theta = 0$, so again $\Lambda$ is pure real.

When $\cos\theta = +1$, (6-97), (6-98) and (6-99) again give $\Lambda = P_1$, so that from (6-105)

$$\Lambda = 2k + \Delta t(L_1 + K_1 L_1 + G_3)$$

The critical case is clearly $K_1 = +1$ so that, using (6-59)

$$\Lambda = 2k + 2k\Delta t G_2$$

From (6-80)

$$\lambda = \frac{2k(1 + G_2\Delta t)}{(2k + G_3\Delta t)(1 + G_2\Delta t)}$$

$$= \frac{2k}{2k + G_3\Delta t} \qquad (6\text{-}114)$$

Thus this maximum of $\lambda$ is less than unity for all $\Delta t$.

When $\cos\theta = -1$, from (6-97) for $|F| \leqslant 1$,

$$\Lambda = 2k + \Delta t(L_1+G_3+K_1L_1) - 2Z(\Gamma_1+K_1L_1\Delta t) - 2W(K_2\Gamma_2+L_2\Delta t) \qquad (6\text{-}115)$$

$$= 2k - 2Z\Gamma_1 - 2WK_2\Gamma_2 - \Delta t\left[L_1(K_1(2Z-1) - 1) - G_3 + 2WL_2\right]$$

As discussed in Section 6.12 this $\Lambda$ becomes critical only when negative, so the critical value of $K_1$ is +1 provided 2Z-1 is positive, and the critical value of $K_2$ is $\pm 1$, taking the same sign as W (i.e. V). Note that

$$\left|K_2\Gamma_2 + L_2\Delta t\right| \leqslant \Gamma_1 + L_1\Delta t$$

from (6-75) and (6-76). Thus we may always conservatively replace $K_2\Gamma_2 + L_2\Delta t$ by $\Gamma_1 + L_1\Delta t$ to give, from (6-115)

$$\Lambda = 2k + \Delta t(2L_1 + G_3) - 2Z(1 + |F|)(\Gamma_1 + L_1\Delta t) \qquad (6\text{-}116)$$

From (6-98) for $F > 1$

$$\Lambda = 2k + \Delta t(L_1+G_3+K_1L_1) - 2W(\Gamma_1+K_1L_1\Delta t) - 2Z(K_2\Gamma_2+L_2\Delta t)$$

$$= 2k - 2W\Gamma_1 - 2ZK_2\Gamma_2 - \Delta t\left[L_1(K_1(2W-1) - 1) - G_3 + 2ZL_2\right]$$

Because $F > 1$, $W > Z$ so again $\Lambda$ becomes critical only when negative. Thus the critical value of $K_1$ is again +1 provided 2W-1 is positive, and again, replacing $K_2\Gamma_2 + L_2\Delta t$ by $\Gamma_1 + L_1\Delta t$ gives (6-116).

From (6-99) with $-F > 1$

$$\Lambda = 2k + \Delta t(L_1+G_3+K_1L_1) + 2W(\Gamma_1+K_1L_1\Delta t) + 2Z(K_2\Gamma_2+L_2\Delta t)$$

$$= 2k + 2W\Gamma_1 + 2ZK_2\Gamma_2 - \Delta t\left[L_1(K_1(-2W-1) - 1) - G_3 - 2ZL_2\right]$$

Because $-F > 1$, $-W > Z$ so we again examine only negative $\Lambda$. The critical value of $K_1$ is again +1 if -2W-1 is positive, and this time we conservatively replace $K_2\Gamma_2 + L_2\Delta t$ by $-\Gamma_1 - L_1\Delta t$, so that again we find (6-116) is the expression for the critical value of $\Lambda$. Now (6-116) can be expressed, using (6-59) and (6-74)

$$\Lambda = 2k + 2kG_2\Delta t - 4k(Z + |W|)(1 + G_2\Delta t)$$

Therefore from (6-80)

$$\lambda = \frac{1 - 2(Z + |W|)}{1 + \frac{G_3\Delta t}{2k}} \tag{6-117}$$

and this is the maximum negative magnitude of $\lambda$ for all F and $\Delta t$. Thus the stability condition (4-56) becomes, for this Semi-Explicit scheme, using (6-9)

$$(|V| + c)\frac{\Delta t}{\Delta s} - \frac{2k-1}{2k}\gamma_2\Delta t \leqslant 1 \tag{6-118}$$

$\gamma_2\Delta t$ is normally small, so we can make the conservative assumption that $\gamma_2\Delta t$ is negligible to give the final stability condition

$$(|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \tag{6-119}$$

Thus the Semi-Explicit scheme depends for stability only on the Courant condition for all $\Delta t$, whereas the "theoretical stability" claimed for other rectangular schemes may exist only as $\Delta t$ tends to zero.

We have assumed that $2Z - 1$, $2W - 1$ and $-2W - 1$ are positive if $|F| < 1$, $F > 1$ and $-F > 1$ respectively.

From (6-119) $\quad (|F| + 1)Z \leqslant 1$

$$Z \leqslant \frac{1}{1 + |F|}$$

$$|W| \leqslant \frac{|F|}{1 + |F|}$$

Thus it is not inconsistent to assume for $|F| \leqslant 1$

$$Z \geqslant \tfrac{1}{2} \tag{6-120}$$

and for $|F| \geqslant 1$

$$W \geqslant \tfrac{1}{2} \tag{6-121}$$

If $|F| = 1$ the equalities in (6-120) and (6-121) must apply for (6-119) to be satisfied, but the more $|F|$ departs from 1 the more latitude there is between (6-119) and (6-120) or (6-121). As shown in Section 4.9 it should be possible to choose $\Delta t/\Delta s$ such that the Courant condition is just satisfied, and (6-120) and (6-121) will then be satisfied. Even though our proof is rigorous only if (6-120) and (6-121) hold, it seems hardly likely that a reduction in the magnitude of W and Z will increase the magnitude of $\lambda$ to the point where instability is possible.

6.15 Comparison of the Results of Analyses

We have used our general expressions (6-79) and (6-80) for the eigenvalues $\Lambda$ and $\lambda$ to apply the sufficient stability condition (4-56) to four difference schemes. We have obtained as sufficient conditions for stability (in addition to conditions such as (6-87) which we have earlier shown to apply to all schemes) the two independent conditions (6-86) and (6-88) for the staggered scheme, condition (6-113) for the CIR rectangular scheme and (6-119) for the Semi-Explicit scheme. We have also proved from (6-90) that the "Unstable" scheme can satisfy stability requirements only if lateral inflow is present and an appreciable time step is used.

All four schemes investigated require the Courant condition (see Section 3.5) as necessary for stability, but for the Semi-Explicit scheme alone the Courant condition is also sufficient. The staggered scheme has an additional stability condition (6-88) on $\Delta t$ which is independent of the Courant condition, while the CIR rectangular scheme combines a condition on $\Delta t$ with the Courant condition, restricting $\Delta t$ to small values which are inefficient for most purposes. This reinforces the conclusion reached in Section 6.4 that a stability condition on $\Delta t$ is introduced unless at least as much weight is put on the forward time line as on the backward time line in the evaluation of the non-homogeneous terms, as of the four schemes tested only the Semi-Explicit scheme uses the forward time line $\mathbf{t} = t + \Delta t$ in the evaluation of such terms.

The "Unstable" scheme justifies its name, requiring such special circumstances for a stable solution that it is unreliable and should not be used.

We compare these schemes further in Chapter 7.

# CHAPTER 7

## THE CHOICE OF A DIFFERENCE SCHEME

### 7.1 Regular and Characteristic Nets

Hyperbolic systems of equations have been solved by difference schemes on two distinct types of finite difference net. The net has either been regular, when the solution was obtained at the intersections of lines parallel to the s and t axes on the s-t plane, or else characteristic, when the solution was obtained at the intersections of selected forward and backward characteristics.

Methods using a characteristic net are the more theoretically fundamental as they follow any small disturbances along their propagation paths and therefore tend to concentrate attention on areas of rapid change in the solution.

On the other hand methods using a regular net are the more practically fundamental in irregular channels as they trace the solution at fixed cross-sections and therefore tend to concentrate attention on areas of the solution where the channel geometry and bed roughness has been studied and hence where the interrelationship between the flow parameters can be specified most accurately. In addition, as indicated by Henderson (1966) p.368, the main influence of a disturbance generally travels as a kinematic wave at a velocity different from those of small disturbances.

Thus the special properties of a characteristic net method are likely to be of advantage only in solutions in which the concentration of characteristics is of primary importance, in particular if travelling "strong discontinuities" (Abbott (1966) p.163), i.e. surges or bores, must be followed along the channel. In this case a characteristic net difference scheme, such as that presented by Liggett and Woolhiser (1967), applied in the manner described by Abbott (1966) Section 4.4, should be more satisfactory than a regular net solution. However, travelling surges are uncommon in natural overland flows, while in artificially controlled flows they are normally avoided, for fear of damaging effects, in subcritical flows by restricting the rate of opening and closing of control gates, and in abrupt transitions from supercritical to subcritical flow by holding the jump stationary in a specially designed structure such as those discussed by Henderson (1966) Section 6.7. Therefore the distinctive properties of a characteristic net method are rarely required. Further, in most solutions the main aim is a description of variations with time of depths and discharges at certain sections along the channel, or a comparison between flows at any time resulting from alternative conditions or operations of controls in the channel. With a regular mesh such requirements are met directly, whereas some interpolation procedure is necessary to produce the required information from a characteristic net solution. As well as demanding considerable programming effort and computer storage, such a procedure must introduce errors, for instance, the products of linearly interpolated velocities and areas must differ from corresponding linearly interpolated discharges; so that the results, even for uniform channels, may be less accurate than those from a stable regular net method.

Thus in all but a minority of cases characteristic net methods require more programming effort, more computer storage and probably more computing effort than a simple, direct regular net method, while offering little expectation of compensating increases in accuracy. It is therefore worth persevering with the development of stable regular net schemes.

## 7.2 The Choice of a Regular Net Difference Scheme

Regular net difference schemes fall into two categories - explicit and implicit schemes. Explicit schemes formulate the difference equations to include only quantities associated with known initial solutions on one or more time lines (lines along which t is constant), plus two dependent variables at a single point on the following time line. Thus the two equations can be rearranged so as to each relate a single dependent variable to known quantities. These equations are usually linear in the unknown, but in the "Semi-Explicit" scheme described in Section 6.14 they may be quadratics. This is further discussed in the next section. The simplest explicit type are one cycle, two level schemes of which perhaps the most direct are the four analysed fully in Chapter 6. Of these the Staggered scheme does not necessarily satisfy consistency, as may be seen from Section 3.6, the "Unstable" scheme cannot be convergent (Section 6.11), and the CIR rectangular scheme is subject to severe stability restrictions on $\Delta t$ (Section 6.13). In contrast the Semi-Explicit scheme is not only stable subject to the Courant condition alone (Section 6.14), but also satisfies consistency as may readily be seen by applying the Taylor series expansion outlined in Section 3.6. Hence the Semi-Explicit scheme can be expected to be convergent for all reasonable mesh widths whose ratio satisfies the Courant condition, and there seems little reason to discard it in favour of any other explicit scheme, as all appear to be governed by at least the Courant condition. Indeed it is difficult to imagine a more straightforward scheme with comparable stability properties.

Implicit schemes simultaneously solve a set of equations connecting the known solutions in one time line to the unknown solutions in the following (forward) time line. Such schemes therefore take the entire solution at any time plus the boundary conditions over the subsequent time step as their domain of dependence, so that as $\Delta t$ tends to zero stability is not related to the value of $\Delta s$ used. This is proved rigorously for one such scheme in Section 5.8, and is often taken to mean that $\Delta t$ and $\Delta s$ can be varied quite independently as each grows large. However, Section 6.4 shows that the non-homogeneous terms raise exactly the same difficulties in implicit schemes as in explicit schemes.

Even if an implicit scheme which formulates the non-homogeneous term with sufficient weight on the forward time line can be proved to be stable for all reasonable mesh widths, as seems possible, maximum solution accuracy for a given computing effort seems unlikely if the mesh width ratio departs markedly from the "natural" ratio governed by the slope of the characteristics. In other words, we would expect more accuracy from an implicit solution using a "natural" mesh width ratio than from one using half the number of time steps and double the number of space steps.

This is confirmed by the experience of Abbott and Ionescu (1967). Their experiments as well as those of Vreugdenhil (1968) showed that large distortions could occur in implicit solutions as the mesh width ratio was increased even though the solutions apparently remained stable. These distortions can be attributed to discretization errors at the boundaries as well as in the solution. As the warning of numerical difficulties given by oscillatory instability is suppressed in implicit schemes, it appears that solutions with implicit schemes using large mesh widths are untrustworthy, particularly if the time step is significantly larger than the characteristic time step.

It is therefore unlikely that any less mesh points can be used with an implicit scheme than with the Semi-Explicit scheme to obtain a solution of given accuracy over a certain area of the s-t plane, while the computation per mesh point should be substantially less with the relatively simple Semi-Explicit scheme.

Thus the Semi-Explicit scheme has no obvious disadvantages compared with implicit schemes for most solutions, while it has the clear advantages of simplicity and directness which are likely to save programming time and computing effort. Implicit schemes may possibly be preferable if the flow or $\Delta s$ vary so widely that the Courant condition leads to absurdly small values of $\Delta t$ over large parts of the solution, or if the flow is near critical in much of the solution, but for most problems the simplicity and stability properties of the Semi-Explicit scheme cannot be bettered by any other difference scheme which we have discussed.

## 7.3 Practical Formulation of the Semi-Explicit Scheme

We now show more specifically how the Semi-Explicit scheme is formulated.

We return to the characteristic form of the Overland Flow equations, (3-25) and (3-26), which can be expressed, writing $\overset{+}{V}_s$, $\overset{+}{w}_s$ in (3-25), $\overset{-}{V}_s$ and $\overset{-}{w}_s$ in (3-26), and using (2-53)

$$V_t + w_t + (V+c)(\overset{+}{V}_s + \overset{+}{w}_s) = \frac{gz_s}{\eta} - \frac{g}{\eta}\left[\frac{V}{DR^m}\right]^{\frac{1}{J}} - \frac{qU}{A\eta} + (q-VH_s)\frac{c}{A} \qquad (7\text{-}1)$$

$$V_t - w_t + (V-c)(\overset{-}{V}_s - \overset{-}{w}_s) = \frac{gz_s}{\eta} - \frac{g}{\eta}\left[\frac{V}{DR^m}\right]^{\frac{1}{J}} - \frac{qU}{A\eta} - (q-VH_s)\frac{c}{A} \qquad (7\text{-}2)$$

First we subtract, then add (7-2) and (7-1) and divide by two to give

$$w_t + \frac{V}{2}(\overset{+}{V}_s - \overset{-}{V}_s + \overset{+}{w}_s + \overset{-}{w}_s) + \frac{c}{2}(\overset{+}{V}_s + \overset{-}{V}_s + \overset{+}{w}_s - \overset{-}{w}_s) = (q - VH_s)\frac{c}{A} \quad (7\text{-}3)$$

$$V_t + \frac{V}{2}(\overset{+}{V}_s + \overset{-}{V}_s + \overset{+}{w}_s - \overset{-}{w}_s)$$

$$+ \frac{c}{2}(\overset{+}{V}_s - \overset{-}{V}_s + \overset{+}{w}_s + \overset{-}{w}_s) = \frac{gz_s}{\eta} - \frac{g}{\eta}\left[\frac{V}{DR^m}\right]^{\frac{1}{j}} - \frac{qU}{A\eta} \tag{7-4}$$

As in the CIR rectangular scheme, we set

$$w_t = \frac{w_a^{b+1} - w_a^b}{\Delta t} \qquad V_t = \frac{V_a^{b+1} - V_a^b}{\Delta t} \tag{7-5}$$

$$\overset{+}{w}_s = \frac{w_a^b - w_{a-1}^b}{\Delta s} \qquad \overset{+}{V}_s = \frac{V_a^b - V_{a-1}^b}{\Delta s} \tag{7-6}$$

$$\overset{-}{w}_s = \frac{w_{a+1}^b - w_a^b}{\Delta s} \qquad \overset{-}{V}_s = \frac{V_{a+1}^b - V_a^b}{\Delta s} \tag{7-7}$$

We are using the notation introduced in Appendix A, and (7-5), (7-6) and (7-7) can be compared with (A-3), (A-4) and (A-5) respectively. Note the space differences are reformulated as indicated in Section 6.12 if $|F| > 1$. Now using (5-5) we can say

$$w_t = 2kc_t \tag{7-8}$$

where k is the value of By/A appropriate to the point $a\Delta s$ over the time between $b\Delta t$ and $(b+1)\Delta t$. k is locally constant at a section for a quasi-simple channel as defined in Section 5.3, and can therefore be evaluated from the known solution at $(a\Delta s,\ b\Delta t)$. Thus we can replace (7-5) by

$$w_t = 2k\frac{(c_a^{b+1} - c_a^b)}{\Delta t} \qquad V_t = \frac{V_a^{b+1} - V_a^b}{\Delta t} \tag{7-9}$$

We can also introduce (cf. (A-8) and (A-9))

$$D_V = \frac{V_{a+1}^b + V_{a-1}^b}{4k} + \frac{2w_a^b - w_{a+1}^b - w_{a-1}^b}{4k} \tag{7-10}$$

$$D_c = \frac{w_{a+1}^b - w_{a-1}^b}{4k} + \frac{2V_a^b - V_{a+1}^b - V_{a-1}^b}{4k} \tag{7-11}$$

where $D_V$ and $D_c$ are clearly functions of the known solution on the time line $t = b\Delta t$ only. Equations (7-3) and (7-4) become respectively, using (7-6) to (7-11)

$$c_a^{b+1} - c_a^b + VD_c \frac{\Delta t}{\Delta s} + cD_V \frac{\Delta t}{\Delta s} = (q - VH_s)\frac{c\Delta t}{2kA} \tag{7-12}$$

$$V_a^{b+1} - V_a^b + 2kVD_V \frac{\Delta t}{\Delta s} + 2kcD_c \frac{\Delta t}{\Delta s} = \frac{gz_s\Delta t}{\eta} - \frac{g\Delta t}{\eta}\left[\frac{V}{DR^m}\right]^{\frac{1}{j}} - \frac{qU\Delta t}{A\eta} \tag{7-13}$$

Now (7-12) is an equation containing one unknown $c_a^{b+1}$ and the variables V, c, $H_s$ and A which we have yet to allocate to a net point. We assume that $D_c$ and $H_s$ are first order quantities as used in Sections 5.6 and 6.5 respectively, so that the formulation of V in (7-12) has little effect on stability. We can therefore evaluate V in (7-12) at the convenient point $(a\Delta s, b\Delta t)$ where V is known. Similarly we can evaluate $H_s$ along the line $t = b\Delta t$. As the lateral inflow is taken to be specified independently, we can now take q as representing the average lateral inflow/unit length at $s = a\Delta s$ over the time step $\Delta t$.

We must evaluate c/A at $(a\Delta s, (b+1)\Delta t)$ in the Semi-Explicit scheme so that $\alpha_1 = 1$, $\alpha_2 = 0$, but there are still two possibilities available. Firstly we can evaluate c/A using (5-21) with $c_a^{b+1}$, $A_a^{b+1}$ in place of $c_a^b$, $A_a^b$ and $c_a^b$, $A_a^b$ in place of c, A respectively. Because the points $(a\Delta s, (b+1)\Delta t)$ and $(a\Delta s, b\Delta t)$ both lie on the line $s = a\Delta s$, $\mathring{f}(s) = 0$. Thus (5-21) gives

$$\frac{c_a^{b+1}}{A_a^{b+1}} \approx \frac{c_a^b}{A_a^b}\left[1 + \frac{c_a^{b+1} - c_a^b}{c_a^b}(1 - 2k)\right] \tag{7-14}$$

Because $D_V$ is a first order quantity it matters little where its coefficient c is evaluated, but as there is no extra difficulty in evaluating this coefficient at $(a\Delta s, (b+1)\Delta t)$ we shall do so as this appears likely to increase stability. Using (7-14) in (7-12) we therefore have

$$c_a^{b+1} = \frac{c_a^b - V_a^b D_c \dfrac{\Delta t}{\Delta s} + \dfrac{c_a^b}{A_a^b}(q - V_a^b (H_s)_a^b)\Delta t}{1 + D_V \dfrac{\Delta t}{\Delta s} + \dfrac{2k-1}{2kA_a^b}(q - V_a^b (H_s)_a^b)\Delta t} \tag{7-15}$$

Alternatively, from (3-14)

$$\frac{c}{A} = \frac{g \cos \phi}{\eta Bc}$$

Hence

$$\frac{c_a^{b+1}}{A_a^{b+1}} = \frac{g \cos \phi}{\eta B_a^{b+1} c_a^{b+1}} \approx \frac{g \cos \phi}{\eta B_a^b c_a^{b+1}} \tag{7-16}$$

Using (7-16) in (7-12) and multiplying by $c_a^{b+1}$ gives

$$(1+D_V \frac{\Delta t}{\Delta s})(c_a^{b+1})^2 + (-c_a^b + V_a^b D_c \frac{\Delta t}{\Delta s})c_a^{b+1} - (q - V_a^b (H_s)_a^b)\frac{\Delta t g \cos \phi}{2k\eta B_a^b} = 0 \tag{7-17}$$

This quadratic has a single valid solution

$$c_a^{b+1} = \frac{\frac{1}{2}(c_a^b - V_a^b D_c \frac{\Delta t}{\Delta s}) + \sqrt{\frac{1}{4}(c_a^b - V_a^b D_c \frac{\Delta t}{\Delta s})^2 + (q - V_a^b (H_s)_a^b)\frac{\Delta t g \cos \phi}{2k\eta B_a^b}(1 + D_V \frac{\Delta t}{\Delta s})}}{1 + D_V \frac{\Delta t}{\Delta s}} \tag{7-18}$$

The other root of the quadratic can be rejected because it must always give a physically impossible negative depth (value of $c_a^{b+1}$). This is because $1 + D_V\Delta t/\Delta s$ and $q - V_a^b(H_s)_a^b$ may be taken as always positive because under the Courant condition $1 - (|V|+c)\Delta t/\Delta s$ is positive and $|D_V| < |V| + c$ for any stable solution (see (7-10)), while we have already assumed that q is positive (Section 6.4) and that $V_a^b(H_s)_a^b$ is small.

Now we have used the assumption that $(c_a^{b+1} - c_a^b)/c_a^b$ is small to obtain (7-14), so that (7-15) is valid while the variations in c are an order of magnitude less than c. We used the assumption that B is approximately constant to obtain (7-16), so that (7-18) is applicable even if the variations in c approach c in magnitude. Thus (7-18) can be applied to the case of lateral inflow into an initially dry flat bottomed channel whereas (7-15) cannot. Hence in general if the lateral inflow is of major importance (7-18) should be used, whereas if the lateral inflow is small, especially if $q < V_a^b(H_s)_a^b$, (7-15) should be used. If $q - V_a^b(H_s)_a^b$ is negligible, of course, both (7-15) and (7-18) reduce to the same linear expression:

$$c_a^{b+1} = \frac{c_a^b - V_a^b D_c \dfrac{\Delta t}{\Delta s}}{1 + D_V \dfrac{\Delta t}{\Delta s}} \qquad (7\text{-}19)$$

Having established $c_a^{b+1}$ we can find $R_a^{b+1}$ and $A_a^{b+1}$ from one to one R/c and A/c relationships which may be defined for any fixed cross-section. These relationships will be explicit in simple channels but may need to be tabulated in advance for irregular channels. $R_a^{b+1}$ and $A_a^{b+1}$ may then be used in (7-13). We also evaluate V at $(a\Delta s, (b+1)\Delta t)$ as required by the Semi-Explicit scheme, and using U = V-u as in (2-20), (7-13) becomes

$$\frac{g\Delta t\ (V_a^{b+1})^{1/j}}{\eta D^{1/j}(R_a^{b+1})^{m/j}} + \left(1 + 2kD_V\frac{\Delta t}{\Delta s} + \frac{q\Delta t}{\eta A_a^{b+1}}\right)V_a^{b+1}$$

$$- V_a^b + 2kc_a^{b+1}D_c\frac{\Delta t}{\Delta s} - \frac{gz_s\Delta t}{\eta} - \frac{qu_a^{b+1}\Delta t}{\eta A_a^{b+1}} = 0 \tag{7-20}$$

If the flow is laminar, j = 1 so that (7-20) is better expressed

$$V_a^{b+1} = \frac{V_a^b - 2kc_a^{b+1}D_c\frac{\Delta t}{\Delta s} + \frac{gz_s\Delta t}{\eta} + \frac{qu_a^{b+1}\Delta t}{\eta A_a^{b+1}}}{1 + 2kD_V\frac{\Delta t}{\Delta s} + \frac{q\Delta t}{\eta A_a^{b+1}} + \frac{g\Delta t}{\eta D(R_a^{b+1})^m}} \tag{7-21}$$

If the flow is turbulent, $j = \frac{1}{2}$, and we can say

$$\left[\frac{g\Delta t\left|V_a^{b+1}\right|}{\eta D^2(R_a^{b+1})^{2m}} + 1 + 2kD_V\frac{\Delta t}{\Delta s} + \frac{q\Delta t}{\eta A_a^{b+1}}\right]V_a^{b+1}$$

$$= V_a^b - 2kc_a^{b+1}D_c\frac{\Delta t}{\Delta s} + \frac{gz_s\Delta t}{\eta} + \frac{qu_a^{b+1}\Delta t}{\eta A_a^{b+1}} \tag{7-22}$$

because $S_f$ takes the same sign as V. We can take it from the Courant condition that $1 + 2kD_V\Delta t/\Delta s$ is positive, so that the coefficient of $V_a^{b+1}$ in (7-22) must always be positive. This means that $V_a^{b+1}$ takes the same sign as the R.H.S. of (7-22), as does $(S_f)_a^{b+1}$. We introduce for convenience

$$K_a = \frac{g\Delta t}{\eta D^2 (R_a^{b+1})^{2m}} \tag{7-23}$$

$$K_b = 1 + 2kD_V \frac{\Delta t}{\Delta s} + \frac{q\Delta t}{\eta A_a^{b+1}} \tag{7-24}$$

$$K_c = -(V_a^b - 2kc_a^{b+1} D_c \frac{\Delta t}{\Delta s} + \frac{gz_s\Delta t}{\eta} + \frac{qu_a^{b+1}\Delta t}{\eta A_a^{b+1}} \tag{7-25}$$

We can now write (7-22) as

$$\frac{-K_c}{|K_c|} K_a (V_a^{b+1})^2 + K_b V_a^{b+1} + K_c = 0$$

whence

$$V_a^{b+1} = \frac{-\frac{1}{2}K_b + \sqrt{\frac{1}{4}K_b^2 + K_c{}^2 K_a / |K_c|}}{\frac{-K_c}{|K_c|} K_a} \tag{7-26}$$

$K_a$ and $K_b$ are both positive and therefore $K_c{}^2 K_a / |K_c|$ is also positive and the other root of the quadratic must give $V_a^{b+1}$ a value with the same sign as $K_c$. This is incompatible with our previous finding from (7-22) that $V_a^{b+1}$ takes the same sign as $-K_c$ and therefore the other root can be rejected as invalid.

We could produce a linear equation in place of (7-26) if we evaluated $(S_f)_a^{b+1}$ by

$$(S_f)_a^{b+1} = \frac{(V_a^{b+1})^2}{D^2 (R_a^{b+1})^{2m}}$$

$$= \frac{(V_a^b)^2}{D^2(R_a^{b+1})^{2m}} \left[ 1 + \frac{V_a^{b+1} - V_a^b}{V_a^b} \right]^2$$

$$\approx \frac{(V_a^b)^2}{D^2(R_a^{b+1})^{2m}} \left[ 1 + 2\,\frac{V_a^{b+1} - V_a^b}{V_a^b} \right] \qquad (7\text{-}27)$$

Unfortunately this linearisation fails whenever $V_a^{b+1} - V_a^b$ approaches or exceeds $V_a^b$ in magnitude, which must certainly happen if V tends to zero or if flow reverses during the solution. Equation (7-26) is therefore more reliable for general use, although computation of the solution takes a little longer than it would if (7-27) were substituted into (7-20) and the resulting linear expression for $V_a^{b+1}$ used. This increase in computation is generally less than 10%, but might reach 25% for a simple channel and small mesh sizes, when the proportion of computation unrelated to the algorithm used for solving for V, such as that devoted to boundary conditions and updating R, A, y and w from c, would reach a minimum. Hence (7-26) is recommended for general use, as this slight increase in computation does not offset the gain in reliability.

## 7.4 The Evaluation of the Quadratic Solution

A standard square root function will generally be used to evaluate (7-18) and (7-26) on a computer, and this method will always be well conditioned numerically for (7-18) because $\frac{1}{2}(c_a^b - V_a^b D_c \Delta t/\Delta s)$ is always positive under the Courant condition, and therefore the numerator of the R.H.S. of (7-18) will always be the sum of two positive quantities. However $K_b$ is positive in (7-26) and therefore the numerator of the R.H.S. of (7-26) is the difference of two positive quantities, so that ill conditioning arises if $K_c^2 K_a / |K_c|$ is small compared with $\frac{1}{4}K_b^2$. We write

$$\frac{K_c^2 K_a}{|K_c|} = |K_c| K_a = p$$

$$-\tfrac{1}{2}K_b + \sqrt{\tfrac{1}{4}K_b^2 + p} = \tfrac{1}{2}\left[-K_b + \sqrt{K_b^2\left(1 + \frac{4p}{K_b^2}\right)}\right]$$

$$= \tfrac{1}{2}\left[-K_b + K_b\left(1 + \frac{4p}{K_b^2}\right)^{\frac{1}{2}}\right]$$

Now $K_a$ is of $O(\Delta t)$, so that as $\Delta t$ tends to zero, p tends to zero, so that the use of the standard square root function will give increasingly ill conditioned results when we decrease the mesh sizes. Thus our actual numerical solution will not be convergent, because of unduly magnified truncation errors, unless a different algorithm is substituted for the standard square root function as $4p/K_b^2$ tends to zero. A suitable algorithm is developed in Appendix C.

## 7.5 Practical Stability Tests

Part II of this thesis concludes with an experimental verification of the stability criteria derived analytically for the CIR rectangular scheme and for the Semi-Explicit scheme. Both are theoretically stable and convergent as the mesh lengths tend to zero provided the Courant condition is satisfied. The only difference between them is the CIR rectangular scheme evaluating the non-homogeneous term (for simplicity in these tests we retain only one variable non-homogeneous term, the frictional resistance) at the point $(a\Delta s, b\Delta t)$ whereas the Semi-Explicit scheme uses the point $(a\Delta s,(b+1)\Delta t)$. Hence any differences in the experimental stability properties can arise only from the difference between evaluating the non-homogeneous term at the backward and forward time steps. Their comparison is therefore a useful test of the validity of our stability analyses. Steady uniform flows were used throughout so that the values of $\sigma$, V and c necessary for the evaluation of the criteria (6-113) and (6-119), were constant throughout the solution. The boundary conditions specified a constant depth and velocity. Eight space increments were found adequate to obviate the damping effect of such regular boundary conditions.

Stability was regarded as experimentally established if the numerical solution of a steady uniform flow did not develop large oscillations after a number of cycles. Even a stable solution cannot necessarily be expected to maintain an exact steady state, as truncation errors arising from the limited precision of any practical calculation will not usually permit the calculation to cycle around a closed loop. Thus a gradual departure from the exact solution indicates instability only if the rate of departure does not ultimately decrease. If we compare the average variation in a cycle with a test number which gradually increases, only unstable solutions will exhibit variations which consistently exceed this number. As the normal precision of an IBM System/360 is about seven decimal digits, a rate of increase in the test number of $2 \times 10^{-7}$ per cycle was regarded as adequate to cover cumulative truncation errors. This rate of increase was vindicated by experience as moderate changes in the rate of increase were found to have only marginal effects on the results of the stability tests. An initial value of $-10^{-6}$ was chosen for the test number to ensure that the solution went through at least five cycles, and it was intended to permit solutions to continue until either the test number overtook the average variation, indicating stability, or else instability was clearly established by the appearance of negative depths in the solution. Unfortunately in a number of tests the variations in the solutions consistently exceeded the test number up to the precautionary limit of 500 cycles without displaying obvious signs of instability. Because the computing time involved in allowing such solutions to continue past 500 cycles was prohibitive, this meant that an arbitrary choice of the maximum number of cycles for any stability test was necessary. Since the test number usually either overtook the average variation after less than 50 cycles, indicating stability, or failed to overtake the variation within 500 cycles, the maximum number of cycles for each of the full series of stability tests was set at 95. After 95 cycles the test number is $1.8 \times 10^{-5}$ which comfortably exceeds the maximum errors discussed in Appendix C (i.e. in (C-13)). The average variation was also checked and the solution was not accepted as stable unless a decrease in variation magnitude occurred over two successive cycles. Thus if the average variation in a solution was still increasing or still exceeded the test number after 95 cycles the solution was regarded as at least marginally unstable.

Initially it was expected that truncation errors alone would always be sufficient to initiate instability, but in a few cases the solution settled into a closed loop from the first cycle. As this intuitively has similar chances of happening whether or not the difference scheme is analytically unstable it was necessary to perturb the initial steady state by introducing a relative error of $10^{-5}$ to the solution at the central mesh point.

Because the principal purpose of the experiments was a practical test of the stability criteria derived in Chapter 6 the objective stability tests described in this section were used to control a step by step procedure to find the critical time steps at the border between stability and instability in practice, for given values of $\sigma$, V, c and $\Delta$s. These borderline time steps could then be compared with their theoretically predicted values, obtained from the equalities in (6-113) and (6-119) for the CIR rectangular and Semi-Explicit schemes respectively.

## 7.6 Numerical Experiments

Program FLOTS1 was developed by the writer to perform the experimental stability analysis on the lines described in Section 7.5. A full listing of the Fortran IV source is included in Appendix D.

Uniform turbulent flows in a wide rectangular channel without lateral flows were considered for simplicity, but the range of variables considered was intended to cover all such flows normally encountered in practice, together with a reasonable range of likely space increments. There are four parameters $y$, $S_f$, $n$ and $\Delta s$ in a difference solution of such flows that can influence $\Delta t$, where $y$ is depth, $S_f$ is friction slope, $n$ is the "Manning n" and $\Delta s$ (or $\Delta x$ if $S_f$ is small as it is here) is the space increment.

For the CIR rectangular scheme, from (6-113) and (6-4) (taking $\eta = 1$, $j = \frac{1}{2}$, $q = 0$)

$$\frac{gS_f\Delta t}{V} + (|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \tag{7-28}$$

For the Semi-Explicit scheme, from (6-119)

$$(|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \tag{7-29}$$

We can express (7-28) and (7-29) in our four basic parameters plus $\Delta t$ because from (3-14) and (2-40), (2-41)

$$c = \sqrt{gy}$$

$$V = \frac{M}{n} S_f^{\frac{1}{2}} y^{\frac{2}{3}}$$

We are interested in the effects of varying four

## 7.6 Numerical Experiments

Program FLOTS1 was developed by the writer to perform the experimental stability analysis on the lines described in Section 7.5. A full listing of the Fortran IV source is included in Appendix D.

Uniform turbulent flows in a wide rectangular channel without lateral flows were considered for simplicity, but the range of variables considered was intended to cover all such flows normally encountered in practice, together with a reasonable range of likely space increments. There are four parameters y, $S_f$, n and $\Delta s$ in a difference solution of such flows that can influence $\Delta t$, where y is depth, $S_f$ is friction slope, n is the "Manning n" and $\Delta s$ (or $\Delta x$ if $S_f$ is small as it is here) is the space increment.

For the CIR rectangular scheme, from (6-113) and (6-4) (taking $\eta = 1$, $j = \frac{1}{2}$, $q = 0$)

$$\frac{gS_f\Delta t}{V} + (|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \qquad (7\text{-}28)$$

For the Semi-Explicit scheme, from (6-119)

$$(|V| + c)\frac{\Delta t}{\Delta s} \leqslant 1 \qquad (7\text{-}29)$$

We can express (7-28) and (7-29) in our four basic parameters plus $\Delta t$ because from (3-14) and (2-40), (2-41)

$$c = \sqrt{gy}$$

$$V = \frac{M}{n} S_f^{\frac{1}{2}} y^{\frac{2}{3}}$$

We are interested in the effects of varying four dimensionless quantities:

$$\frac{gS_f\Delta t}{V}, \ (V + c)\frac{\Delta t}{\Delta s} \text{ and their ratio } \frac{gS_f\Delta s}{V(V+c)},$$

together with the Froude number F which influences the formulation of the difference schemes. Now as we increase y the natural scale of the problem normally increases and we expect to also increase $\Delta s$ and $\Delta t$. From (4-61) with $\cos\phi = 1$

$$V + c = \frac{F+1}{F^{\frac{1}{3}}} g^{\frac{1}{3}} \left[\frac{VyB}{B}\right]^{\frac{1}{3}}$$

$$= \frac{F+1}{F^{\frac{1}{3}}} g^{\frac{1}{3}} \frac{M^{\frac{1}{3}}}{n^{\frac{1}{3}}} S_f^{1/6} y^{5/9}$$

Thus for given $S_f$ and n, V+c varies roughly as $y^{5/9}$ because $(F+1)F^{-\frac{1}{3}}$ is fairly constant as shown in Chapter 4. We therefore chose to vary $\Delta s$ as $y^{11/9}$ in order to keep $\frac{gS_f\Delta s}{V(V+c)}$ approximately constant for constant $S_f$ and n, and this gave increases of $\Delta s$ with y which intuitively matched $\Delta s$ to the increasing scale of the problem.

The stability tests were organised in a four level hierarchy. The top level varied depth; a typical minimum turbulent depth of 0.04 ft, a typical maximum of 25 ft, and an intervening depth of 1 ft were used. The second level assigned three typical values of $\Delta s$, differing by a factor of $\sqrt{10}$, to each depth, and the third level varied the water surface slope from $10^{-4}$ to $10^{-3}$ to $10^{-2}$. Finally, in subroutine IZETS1 the value of Manning's n (EM) was varied from 0.012 to 0.024 to 0.048.

Thus the value of $\Delta t$ which just caused instability in the two formulations was compared experimentally over 81 combinations of y, $\Delta s$, $S_f$ and n which should cover common difference solutions of turbulent flows in wide rectangular channels.

The time steps $\Delta t$ were initially set using the equalities in (7-28) and (7-29) respectively and if the resulting solutions were stable these time steps were increased by 10%. This process was continued until the solutions become unstable. If the first solutions were unstable the original time steps were decreased by 5% repeatedly until stable solutions were reached. The ratios of the lowest unstable time steps to the original time steps were then stored. The lowest unstable time steps, for given y, $\Delta s$, $S_f$ and n, were taken as marking the experimental border between stability and instability and retained rather than the highest stable time steps, because as discussed in the previous section solutions of marginal stability tended to be regarded as unstable by the objective stability criteria necessary for automatic stability tests.

The function of the seven subprograms in FLOTS1 can now be described briefly (See the listing in Appendix D).

The main subprogram MAIN set up the uniform flow conditions for the range of cases described above and then called the other six subprograms as required. After TESTS1 signalled either stability or instability for each solution with a trial time step, MAIN checked for the stability/instability border and if necessary continued the step by step search as described above. If the stability/instability border was detected, the critical time step at which instability just emerged was stored as a ratio CRP(K,L) with the critical time step predicted from the equality in (7-28) or (7-29).

Subroutine IZETS1 set Manning's n, and initialized several variables such as V and c for each solution.

Subroutine HEDTS1 printed output giving the values of the relevant parameters for each stability test and set $\Delta t$, as well as initializing functions of $\Delta t$. It also introduced the perturbation at the central mesh point.

Subroutine STRTS1 carried a solution forward one time step using the CIR rectangular scheme in the form described by Stoker (1957), p.477.

Subroutine BNTTS1 carried a solution forward one time step using the Semi-Explicit scheme as detailed in Sections 7.3 and 7.4. Note that 2.23 = 0.558 x 4 (see (C-7) in Appendix C) to sufficient accuracy. Note also that although there are 25 statements in BNTTS1 and only 11 in STRTS1 many of those in BNTTS1 are quite simple so the ratio of storage requirements, exclusive of COMMON, is only about 1.4. As the difference solution algorithm normally occupies only a small portion of the total storage requirements of a program, this storage increase is insignificant.

In subroutine TESTS1 the stability tests described in Section 7.5 were performed, and if the tests revealed neither stability nor instability the solution was returned to STRTS1 or BNTTS1 for a further cycle. If the tests revealed either stability or instability the appropriate message and the final solution were printed. The result was then signalled to MAIN for the test for borderline instability covered in the description of MAIN.

Subroutine TABTS1 tabulated a summary of the experimental results based on the critical time steps at which stability just emerged for each difference scheme in the 27 combinations of $\Delta s$, $S_f$ and n associated with each value of depth. These tables appear in Appendix E, together with a sample of the printout from TESTS1.

## 7.7 Discussion of Results

We refer to Appendix E, where the stability test results are summarised in pages E1, E2 and E3 for uniform flows at depths of 0.04 ft, 1.0 ft and 25 ft respectively. On page E3 four lines of results are enclosed in a box and the full output from TESTS1 summarised in those four lines is given in pages E4 and E5. This typical output illustrates the variation with the time step of the stability properties of the difference solutions.

The heading at the top of page E4 gives depth (y), mesh length ($\Delta s$), slope ($S_f$), Manning n, and also the Froude Number, that apply unchanged for each trial $\Delta t$. The "Stoker Finite Difference Scheme" is that employed by STRTS1 and tested first. The initial time step of 163.4 seconds was obtained from the equality in (7-28) and for the given parameters implies

$$(V+c)\frac{\Delta t}{\Delta s} = 0.37$$

$$\frac{gS_f\Delta t}{V} = 0.63$$

Note DX is used for $\Delta x$ which is equivalent to $\Delta s$ when $\cos\phi = 1$ (see Figure 2-3) as we are assuming here. For this time step the trial Criterion Ratio (C.R.) is unity. The Criterion Ratio is defined

$$\text{C.R.} = \frac{\text{Minimum unstable time step from experiments}}{\text{Minimum unstable time step from (7-28) or (7-29)}} \qquad (7\text{-}30)$$

Naturally the denominator of the C.R. is defined from the equality in (7-28) or (7-29) according as the CIR or Semi-Explicit scheme is being tested. With the trial C.R. = 1.00 a stable solution was detected after 9 cycles because the average proportional change of $6.8 \times 10^{-7}$ was overtaken by the test number of $(9 \times 2 \times 10^{-7}) - 10^{-6}$ (see Section 7.5). The average proportional change must also have been decreasing for at least two cycles. The ratio of the depth and friction slope at the central mesh point to the uniform depth and the uniform flow friction slope respectively are also recorded, together with the final solutions of c and V at all mesh points.

Because the initial time step was stable it was increased by 10%, making the trial C.R. equal to 1.10 and again a stable solution was detected. However a second increase in time step resulted in a solution which failed to satisfy the stability criteria within 95 cycles, so it was rejected as unstable in the

96th cycle. (Note that the magnitude of the average proportional change, the ratios of typical depth and friction slope to their uniform value, and the final solution itself all clearly support this rejection in this case.)

Therefore the time step of 196.08 seconds was recorded as the critical time step at which instability just emerged for difference scheme No.1, i.e. the CIR scheme, and was later printed with its associated data in the first line of the box in page E3.

The "Proposed Finite Difference Scheme" (i.e. the Semi-Explicit scheme employed by BNTTS1) was then tested in the same way although the initial time step was much larger because (7-29) is far less restrictive than is (7-28) when $gS_f\Delta t/V$ is large as in this particular test. In this case the solution quite suddenly became unstable when the C.R. lay somewhere between 1.40 and 1.50, so the time step of 659.95 seconds was recorded as critical for difference scheme No.2, i.e. the Semi-Explicit scheme, and was duly printed in the second line of the box in page E3. The column headings on page E3 (and E1, E2) correspond in an obvious manner with the data on page E4 except for the fifth and sixth column heading. These are defined by

$$\frac{DT2}{DT1} = \frac{\text{Experimental critical time step (Semi-Explicit)}}{\text{Experimental critical time step (CIR)}} \tag{7-31}$$

$$\frac{GSDX}{V(V+C)} = \frac{gS_f\Delta s}{V(V+c)} = \frac{gS_f\Delta t}{V}\Big/(V+c)\frac{\Delta t}{\Delta s} \tag{7-32}$$

Thus in this example the Semi-Explicit difference solution just became unstable with a time step 3.37 times as large as that with which the CIR difference solution just became unstable. The ratio in (7-32) is dependent only on the given variables y, $\Delta s$, $S_f$ and n and expresses the relative importance of the two terms on the LHS of (7-28).

The third and fourth lines in the box on page E3 summarise the results on page E5. These are given in full to illustrate typical stability properties of the solutions of uniform flows with a high Froude Number. In this case F = 3.73. After instability was detected in the first trial with the CIR scheme, the time step was reduced by 5% seven times before a stable solution was obtained. Note that the instability became successively less severe with each reduction in time step which is consistent with progressive decreases towards unity in the magnitude of the amplification factor, i.e. the factor by which the initial errors in each cycle are multiplied. In this case of course the second to last trial time step was taken as the critical one.

With the Semi-Explicit scheme the initial solution in this case was rather surprisingly stable. Presumably this means simply that the initial distribution of errors corresponded with a Fourier series with no harmonics with a value of $\theta$ for which the stability criterion was infringed. The second trial solution was only little less stable while even the third was only marginally unstable, with an average proportional change less than three times the test number after 96 cycles. However, irregularities were present in the final solution in the third trial, so that an ultimate breakdown in the solution was probable.

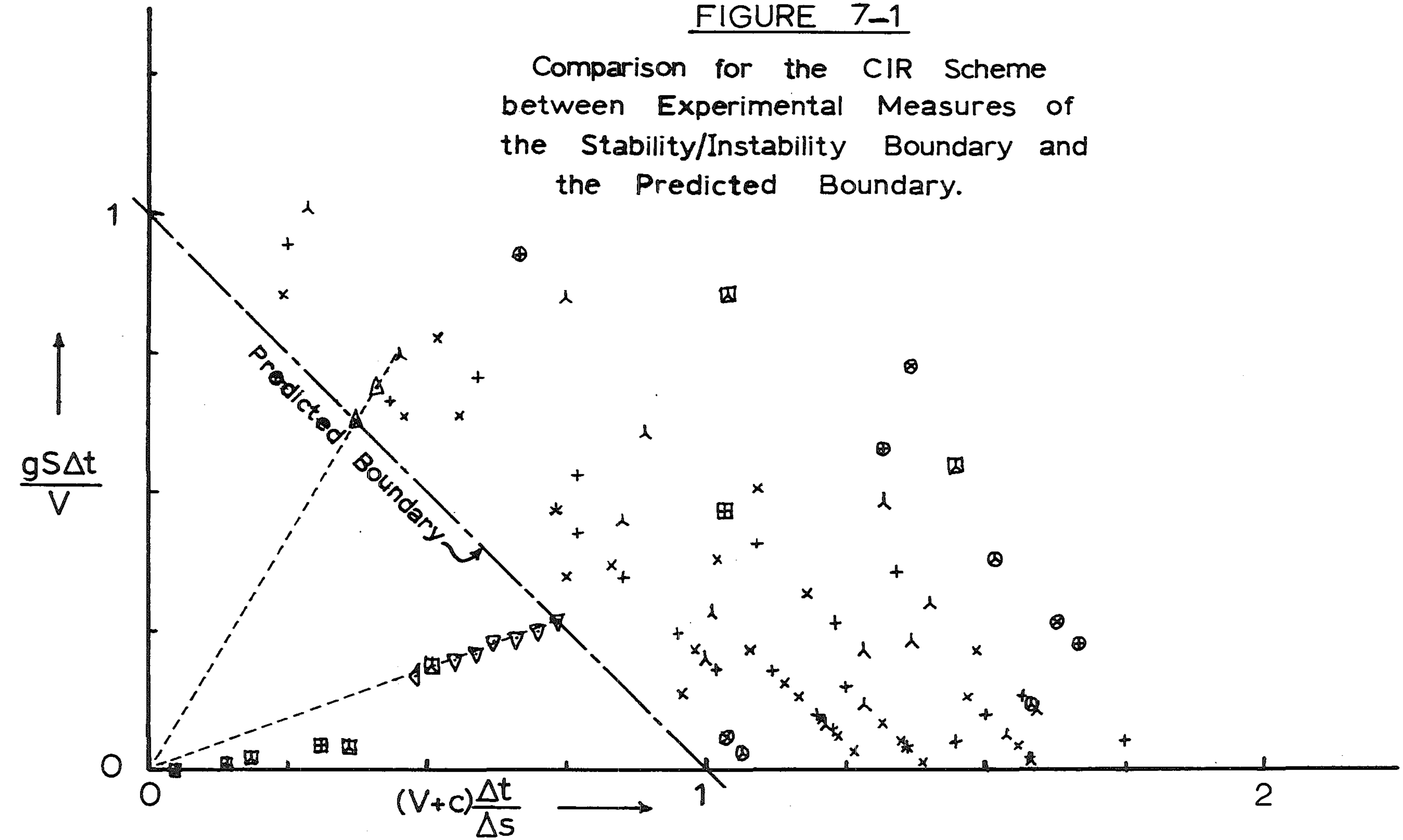
Figures 7-1, 7-2, 7-3 and 7-4 are all plotted from the results tabulated in Appendix E.

Figure 7-1 is a plot of the tenth column against the ninth column in pages E1, E2 and E3 for the CIR scheme and Figure 7-2 is the corresponding plot for the Semi-Explicit scheme. Both the ninth and the tenth column incorporate the critical time step at which instability just emerged, multiplied by a constant for each given set of y, $\Delta s$, $S_f$ and n. Hence each point with coordinates given by columns nine and ten can be joined to the origin by the line, representing variation in $\Delta t$ alone, along which $\Delta t$ was varied in the search for the critical time step. Thus between each plotted point and the origin the time step was stable, while beyond each plotted point on a line through the origin the time step was unstable. For example, in the trials to find each result in the box in page E3, the time step was moved along the lines shown dotted, passing through trial points with coordinates given in pages E4 and E5, until the critical time step was established as plotted.

Now throughout these tests we used the formulation for $|F| \leqslant 1$ in both schemes as this provided a test of the reaction of each scheme to an incorrectly modelled domain of dependence when F exceeded unity.(F was positive in all tests).

# FIGURE 7–1

Comparison for the CIR Scheme between Experimental Measures of the Stability/Instability Boundary and the Predicted Boundary.

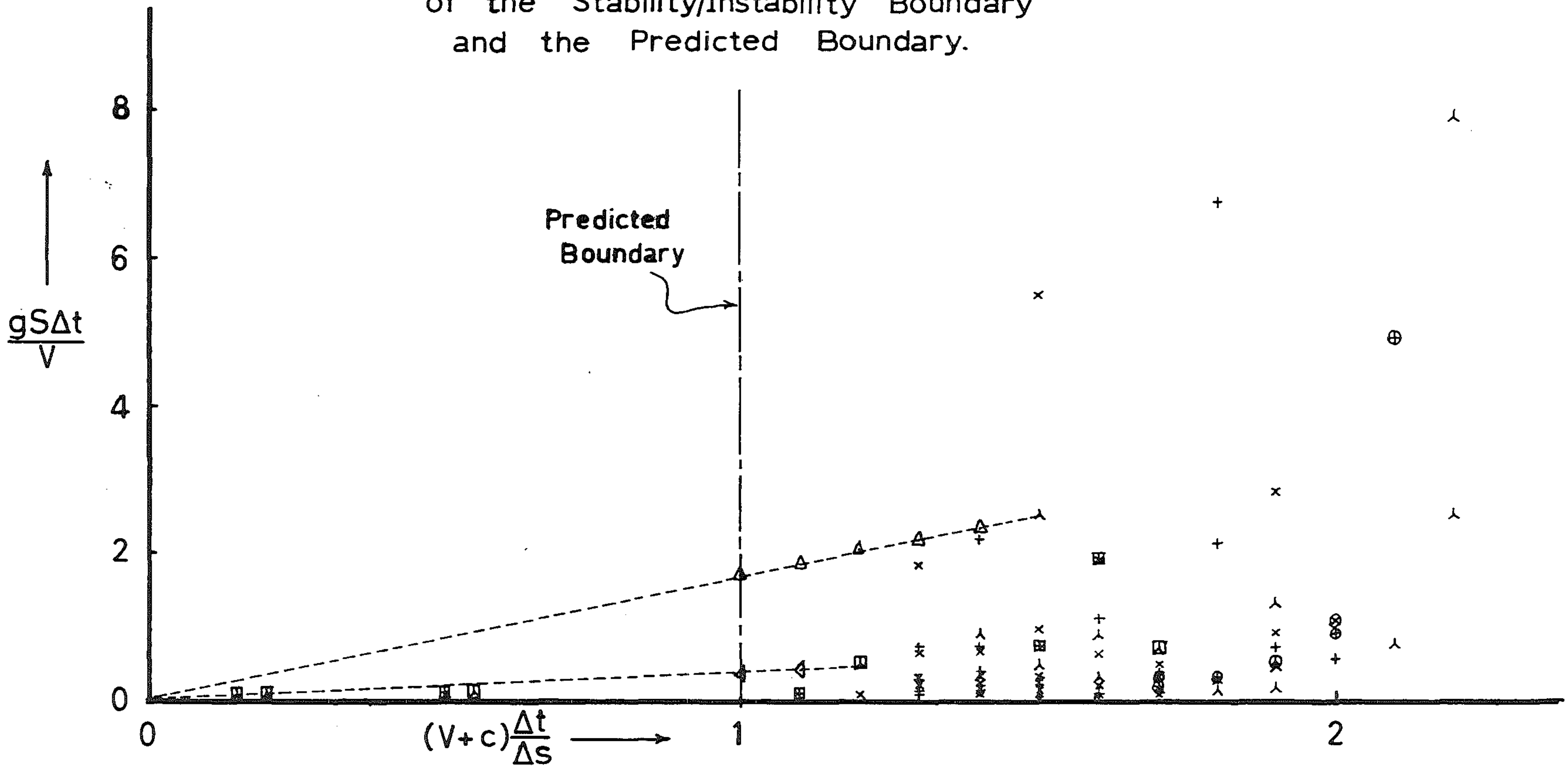



| Points at which Experimental Stability/Instability Boundary Detected | | | |
|---|---|---|---|
| | Depth = 0·04ft | Depth = 1·0ft | Depth = 25ft |
| F < 1 | × | + | ⅄ |
| 1 < F < 1·5 | ⊗ | ⊕ | ⊕ |
| F > 1·5 | ⊠ | ⊞ | ⊠ |

| Sample Trial Points | |
|---|---|
| | Depth = 25ft |
| Stable, F = 0·30 | △ |
| Stable, F = 3·73 | ◁ |
| Unstable, F = 3·73 | ▽ |

# FIGURE 7-2

Comparison for the Semi-Explicit Scheme between Experimental Measures of the Stability/Instability Boundary and the Predicted Boundary.

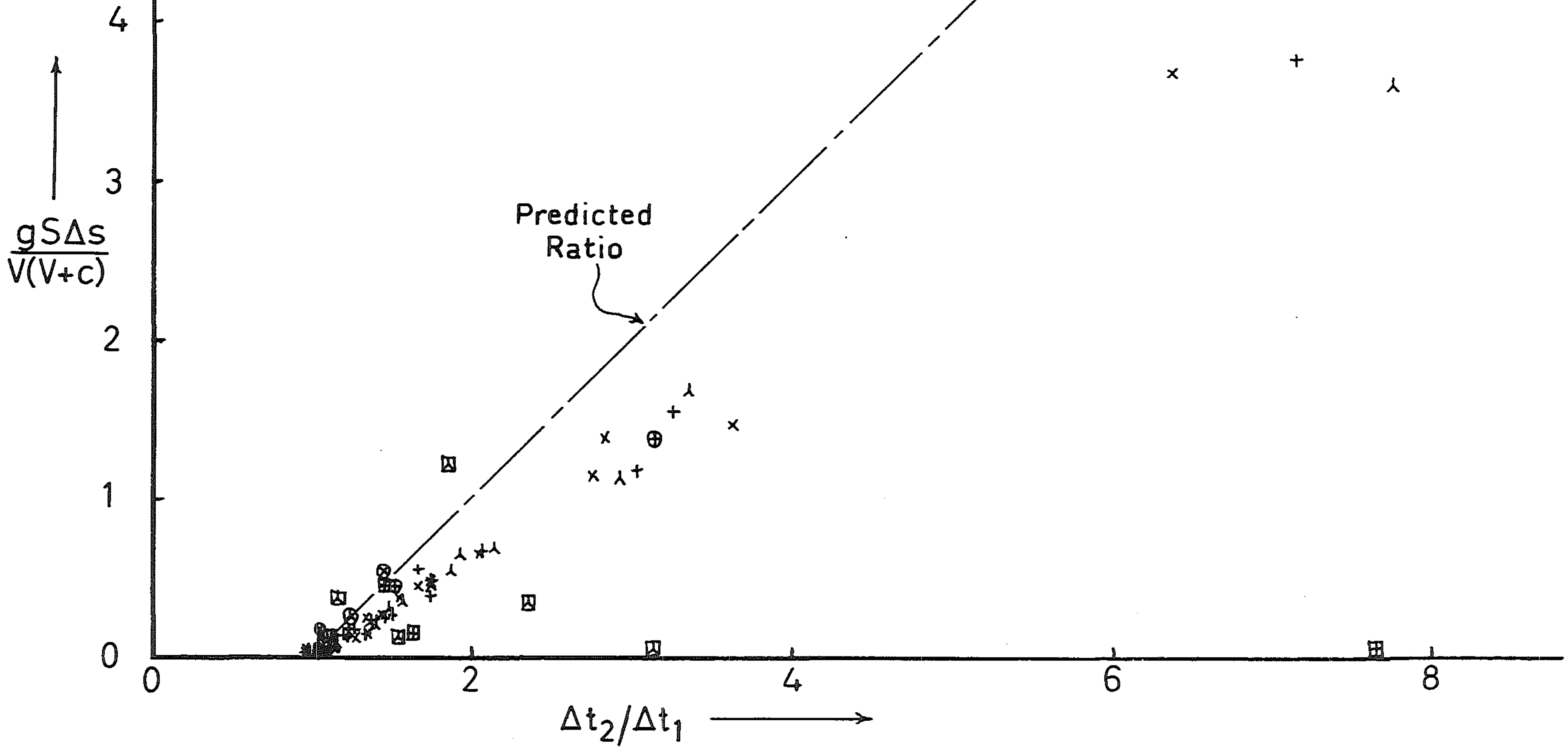



| Points at which Experimental Stability/Instability Boundary Detected | | | |
|---|---|---|---|
| | Depth = 0.04ft | Depth = 1.0ft | Depth = 25ft |
| F < 1 | × | + | ⅄ |
| 1 < F < 1.5 | ⊗ | ⊕ | ☮ |
| F > 1.5 | ⊠ | ⊞ | ⌧ |

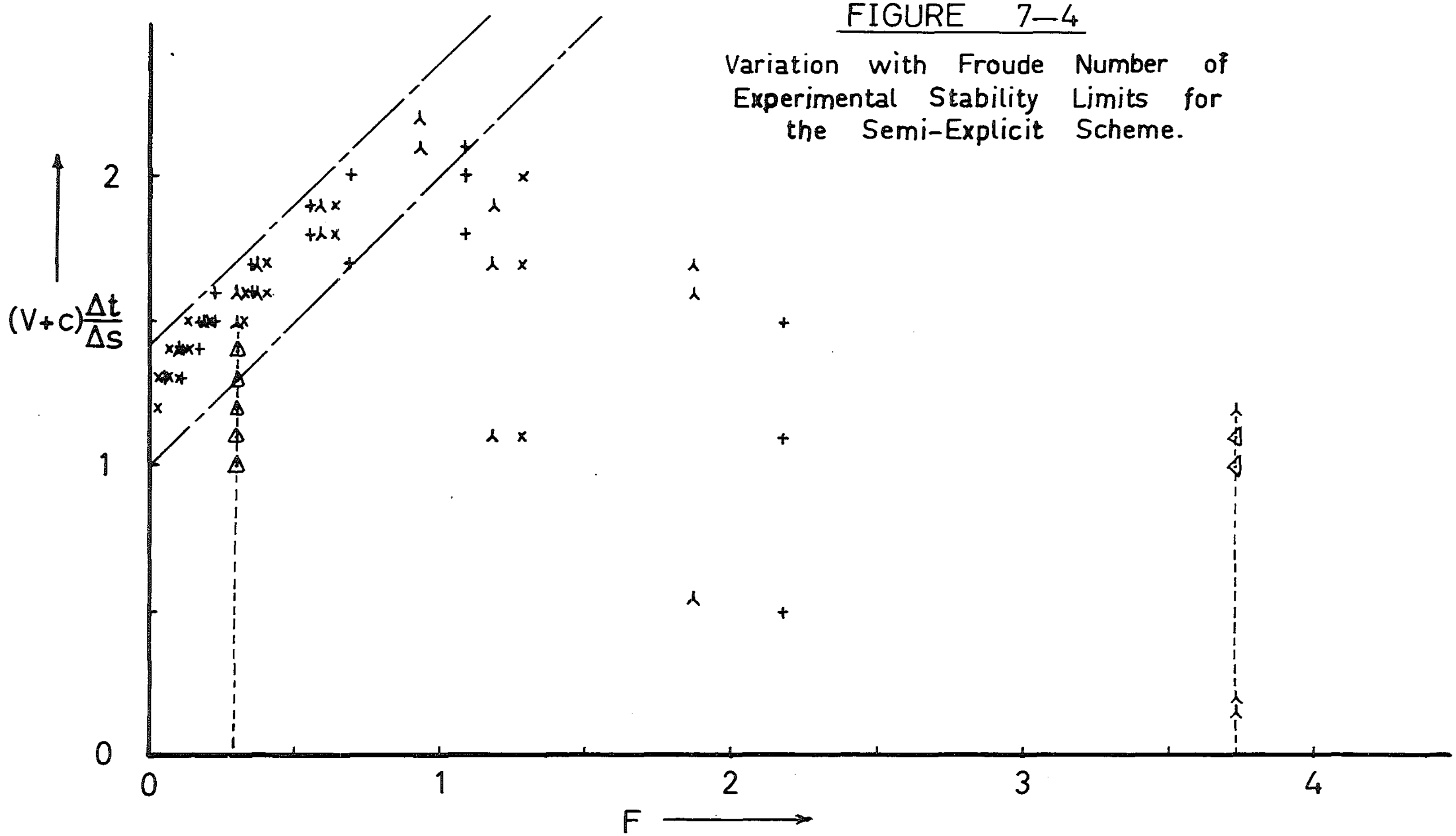
| Sample Trial Points | |
|---|---|
| | Depth = 25ft |
| Stable, F = 0.30 | △ |
| Stable, F = 3.73 | ◁ |

## FIGURE 7–3

The Ratio Between Critical Time Steps for the Semi-Explicit and CIR Schemes.




| Points at which Experimental Stability/Instability Boundary Detected | | | |
|---|---|---|---|
| | Depth = 0.04ft | Depth = 1.0ft | Depth = 25ft |
| $F < 1$ | × | + | ⅄ |
| $1 < F < 1.5$ | ⊗ | ⊕ | ☮ |
| $F > 1.5$ | ⊠ | ⊞ | ⅄⃞ |




# FIGURE 7–4

Variation with Froude Number of Experimental Stability Limits for the Semi-Explicit Scheme.

| Points at which Experimental Stability/Instability Boundary Detected | | |
|---|---|---|
| Depth = 0.04ft | Depth = 1.0ft | Depth = 25ft |
| × | + | ⅄ |

| Sample Trial Points | |
|---|---|
| | Depth = 25ft |
| Stable, F = 0.30 | △ |
| Stable, F = 3.73 | ◁ |

Thus the experimental results can strictly be compared with the analytical stability predictions of Chapter 6 only when $F \leqslant 1$ in the numerical tests, and we have accordingly distinguished all results from solutions with $F > 1$ from those with $F < 1$ in Figures 7-1, 7-2 and 7-3. Our analysis also predicts, from (5-10), unconditional instability for $F > 1.5$ as $m = \frac{2}{3}$ and $k = 1$ in all tests, so that test results from solutions with $F > 1.5$ are also identified in Figures 7-1, 7-2 and 7-3.

For the CIR scheme, from (7-28), our stability analysis predicts that none of the points plotted should lie between the line

$$\frac{gS_f\Delta t}{V} = 1 - (V+c)\frac{\Delta t}{\Delta s}$$

and the axes unless they are associated with solutions with $F > 1$. This boundary line is chain dotted in Figure 7-1 and the above prediction is clearly verified.

Similarly for the Semi-Explicit scheme, from (7-29), our stability analysis predicts that no points should lie to the left of the line

$$(V+c)\frac{\Delta t}{\Delta s} = 1$$

unless they are associated with solutions with $F > 1$. This line is chain dotted in Figure 7-2 and again the analytical prediction is verified.

Now we would expect the analytical boundaries between the stable and unstable regions of Figures 7-1 and 7-2, shown chain dotted, to be conservative as theoretical stability analysis must postulate the worst possible combination of initial errors. We would also expect the degree of conservatism to vary in a largely random manner from test to test according to how closely the initial errors approach the worst combination. By a similar argument we might expect stable solutions to be possible with random values of $\Delta t$ for $F > 1$ and even for $F > 1.5$. All these expectations are shown to be justified by Figures 7-1 and 7-2 because all the points plotted, neglecting the trial points from pages E4 and E5 plotted as triangles, can be regarded as experimental spot estimates of the respective boundaries between the stable and unstable regions of the Figures.

Comparison of Figure 7-2 with Figure 7-1 shows the effect on stability of formulating the non-homogeneous term at the forward rather than backward time line, because many of the experimental spot estimates of the stability/instability boundary lie well to the left of a line

$$(V+c)\frac{\Delta t}{\Delta s} = 1$$

representing the Courant condition, in Figure 7-1 with $F < 1$ while none of the corresponding spot estimates in Figure 7-2 violate this line even though values of $gS_f\Delta t/V$ almost eight times the maximum value possible with the CIR scheme were used successfully.

Thus the respective stability conditions (7-28) and (7-29) are confirmed by the numerical experiments.

If we assume that instability emerges as soon as (7-28) and (7-29) are infringed, the ratio of the critical time steps of the

Semi-Explicit scheme and the CIR rectangular scheme is, from the equalities in (7-28), (7-29)

$$\frac{\Delta t_2}{\Delta t_1} = \frac{\Delta s}{V + c}\left[\frac{V + c}{\Delta s} + \frac{gS_f}{V}\right] = 1 + \frac{gS_f \Delta s}{V(V+c)} \qquad (7\text{-}33)$$

From (7-31) we have in the fifth column of the tabulated results in pages E1, E2 and E3 an experimental measure of the ratio $\Delta t_2/\Delta t_1$, while from (7-32) the corresponding values of $\frac{gS_f \Delta s}{V(V+c)}$ are listed in the sixth column. Figure 7-3 is a plot of the sixth column against the fifth column, and the experimental relationships between these two dimensionless parameters can be compared with (7-33), shown chain dotted. It can be seen that for the great majority of solutions that (7-33) gives an underestimate of the ratio of the critical time steps, and we shall discuss possible reasons for this in a moment.

Now we can regard the ratio of critical time steps as virtually the ratio of the maximum stable time steps, which is of considerable importance in comparing the computation speed of the two schemes. From the experience of the writer with the two schemes, the Semi-Explicit scheme increases the computation time for a single cycle of the solution by a factor varying from less than 1.1 to about 1.3 compared with the CIR rectangular scheme, depending on the proportion of computation devoted to boundary conditions and other matters independent of the difference scheme used. The number of cycles needed to cover a given time period in a solution is inversely proportional to the time step used, so that as long as $\Delta t_2/\Delta t_1$ exceeds the ratio of computation times for a single cycle the Semi-Explicit scheme will be faster to use than the CIR rectangular scheme. We may therefore conclude from Figure 7-3 that the Semi-Explicit scheme offers a great speed advantage over the CIR rectangular scheme for approximate solutions in a coarse mesh without a corresponding disadvantage for accurate solutions on a fine mesh.

Returning to the discrepancy on Figure 7-3 between equation (7-33) and the experimental results, it must be remembered that as $\frac{gS_f\Delta s}{V(V+c)}$ increases, the maximum stable time step for the CIR rectangular scheme departs from the "natural" time step dictated by the slope of the characteristics, so that the CIR rectangular scheme represents the actual behaviour of the characteristics less successfully. Thus as $(V+c)\Delta t/\Delta s$ decreases, the stability condition (7-28) might be expected to become less conservative, and there is some suggestion of this in Figure 7-1. In contrast the Semi-Explicit scheme can employ the "natural" time step in all circumstances, so the stability condition (7-29) might be expected to be equally conservative in all cases. In fact there is a considerable spread of discrepancies between the experimental stability limits and (7-29) in Figure 7-2, but these are evidently not related to the value of $gS_f\Delta t/V$.

Rather the experimental departure from the stability condition (7-29), i.e. the Courant condition, for the Semi-Explicit scheme appears to be related to the Froude Number of the solution.

Figure 7-4, where the ninth column



This is shown in Figure 7-4, where the ninth column of results for the Semi-Explicit scheme, i.e. every second result, is plotted against the fourth column in the tabulation of results in pages E1, E2 and E3. Note the step by step trial values of $(V+c)\Delta t/\Delta s$ from pages E4 and E5 are also shown. If we plot the lines (shown chain dotted)

$$(V+c)\frac{\Delta t}{\Delta s} = 1 + F$$

and

$$(V+c)\frac{\Delta t}{\Delta s} = 1.4 + F$$

we see that all Semi-Explicit scheme tests with $F < 1$ indicated that marginal instability first occurred in the band between these lines. The width of the band can be accounted for by the random nature of the initial errors as already discussed, but the trend revealed by Figure 7-4 is quite systematic.

If we assume locally constant values of c and V, the actual domain of dependence of the point $(a\Delta s,(b+1)\Delta t)$ on the line $t = b\Delta t$ is $2c\Delta t$, while in the Semi-Explicit scheme it is $2(V+c)\Delta t$, assuming that (7-29) is just satisfied. Thus the ratio (numerical domain of dependence)/(actual domain of dependence) is 1+F. Figure 7-4 therefore suggests that the Semi-Explicit scheme will damp out instabilities as long as the domain of dependence is adequately represented in size even if the forward characteristic originates from outside the numerical domain of dependence. The results of the stability tests on the CIR rectangular scheme tend to suggest that a similar effect applies to all regular net schemes, at least for uniform flows. However, it is clear that when $\Delta t$ is chosen prior to a solution, a slight underestimate of the maximum value of V + c is unlikely to have drastic consequences, so the methods of Section 4.9 should permit a suitable value of $\Delta t$ to be chosen with confidence prior to a Semi-Explicit scheme solution.

Similarly, it is clear from Figures 7-1 and 7-2 that if the flow becomes just supercritical in parts of the solution it should not be necessary to redefine the difference mesh to avoid instability, as the only tests for which (7-28) or (7-29) did not hold were those in which F exceeded 1.5 when the flows would be physically unstable in any case. This is not to say of course that a solution of a problem in supercritical flow should be attempted using the formulation for subcritical flow, as the fundamental nature of the solution would be distorted.

We have distinguished between the test results for the three values of depth in Figures 7-1, 7-2, 7-3 and 7-4 to investigate whether the experimental stability properties are in any way dependent on the scale of the physical problem. However no consistent trend related to the variation in depth from 0.04 feet to 25 feet is apparent in any of the four figures. This conforms to expectation.

In summary, it was clearly established that condition (7-28) is the correct stability condition for the CIR rectangular scheme in the form described by Stoker, and that the Courant condition (7-29) applies to the Semi-Explicit scheme. The analysis of Chapters 5 and 6 gave accurate qualitative predictions of the stability properties of the entire range of solutions tested, with the quantitative stability conditions being slightly conservative as expected.

## 7.8 The Choice of A Difference Scheme

In this chapter we have briefly discussed a wide range of difference schemes for use in problems of overland flow, and in Sections 7.1 and 7.2 we showed that there is considerable evidence to support the choice of the Semi-Explicit scheme. Much of this evidence was based on qualitative comparisons, derived from the stability analysis of Chapters 5 and 6, of the stability properties of difference schemes. This theoretical analysis has now been substantiated by numerical experiments, establishing more securely the superiority of the Semi-Explicit scheme. This scheme is the same as the CIR rectangular scheme proposed by Courant, Isaacson and Rees (1952), except that the non-homogeneous terms are evaluated at the forward rather than the backward time line, as detailed in Section 7.3. The scheme is basically explicit but the name "Semi-Explicit" was used because in common with implicit schemes the unknown solution is incorporated in a number of terms.

Some coincidence between the sound stability properties of the Semi-Explicit scheme and the properties of characteristics are evident. The non-homogeneous terms are evaluated at the point $(a\Delta s,(b+1)\Delta t)$, from which, provided the Courant condition is satisfied, characteristics can be traced back to pass through the intervals across which the respective space differences are evaluated. Thus we are solving our equations along the characteristics to a greater extent in the Semi-Explicit scheme than in any other regular net scheme discussed.

Although we have concentrated on stability, and to a lesser extent consistency, in comparing the convergence properties of the simpler regular net difference schemes, three other factors may also in practice affect the behaviour of difference solutions. These are ill conditioning, boundary condition formulation, and round off errors.

We define ill conditioning as an undesirable combination of properties of the solution and of the difference scheme which results in gross magnification of a small error over a single mesh. Examples already cited are the use of (7-15) when $A_a^b$ is small and the use of a standard square root function in (7-26) when $\Delta t$ is small. Means of circumventing these cases of ill conditioning are given in Sections 7.3 and 7.4 respectively and illustrate the point that the expected properties of a solution must be carefully considered in the formulation of a difference scheme algorithm if the solution is to be not only stable but also accurate.

Boundary conditions are of great importance but their properties are sufficiently complex to warrant fuller discussion than is possible herein. However it is fundamental to our local application of Fourier series stability analysis that genuine instability arises only from the amplification of errors from an initial time line and that the influence of boundary conditions can therefore be treated as superimposed on the stability properties of the solution. Thus for a solution to be accurate, both properly formulated boundary conditions and a stable difference scheme are required.

Round off errors are not normally important because most computers should carry sufficient significant digits for solutions of adequate accuracy to be obtained. However if the difference net is refined so much that finite differences of the solution across a single mesh are severely rounded then extra precision is required, in which case the rule derived by Courant, Isaacson and Rees (1952), and quoted at the end of Section 3.5, should be helpful.

While ill conditioning, boundary condition formulation, and round off errors must all be considered in the choice of a difference scheme, there is no evidence that the Semi-Explicit scheme is any more vulnerable to boundary condition difficulties or round off errors than any other scheme, while we have isolated and corrected two possible cases of ill conditioning in the Semi-Explicit scheme in Sections 7.3 and 7.4.

We conclude that for general purpose use in the solution of problems in overland flow equations, the Semi-Explicit scheme offers a combination of simplicity, speed of computation, and predictable stability properties unmatched by any other difference scheme considered.

# PART III - THE DIFFUSION ANALOGY SIMPLIFICATION

## CHAPTER 8

## DIFFUSION ANALOGY THEORY

### 8.1 Introduction

We now discuss the implications of one possible simplification of the Overland Flow equations which leads to a single equation analagous to diffusion equations derived from other branches of physics.

By neglecting the acceleration and extra momentum terms on the L.H.S. of the equation of motion (2-57) we have

$$S_f = z_s - y_s \cos\phi \tag{8-1}$$

Thus $S_f$ can be regarded as independent of V and its derivatives, which means that (2-54) is then an explicit definition of V. Hence we can substitute for V in the continuity equation (2-56) to obtain

$$DR^m S_f^{\,j} \left[ A_s + A\left( m\frac{R_s}{R} + j\frac{(S_f)_s}{S_f} \right) \right] + A_t = q \tag{8-2}$$

For a simple channel (see Section 2.8) this becomes, after rearranging and dividing throughout by the surface width B,

$$y_t + (1 + \frac{m}{k})\, Vy_s + \frac{jVy}{kS_f}\, z_{ss} + \frac{Vy}{k}\, \frac{E_s}{E} - \frac{q}{B} = \frac{jVy\cos\phi}{kS_f}\, y_{ss} \tag{8-3}$$

where k and E are defined in (2-35) and (2-37). Note that we are here taking k and the ratio y/R = L (see (2-36)) to be approximately constant rather than functions of s. We substitute

$$c = (1 + \frac{m}{k})V \tag{8-4}$$

$$K = \frac{jVy \cos \phi}{kS_f} \tag{8-5}$$

Thus (8-3) becomes

$$y_t + cy_s + \frac{Kz_{ss}}{\cos \phi} + \frac{KS_f}{j \cos \phi} \frac{E_s}{E} - \frac{q}{B} = Ky_{ss} \tag{8-6}$$

If the channel is prismatic with no lateral inflow $z_{ss} = E_s = q = 0$, and we have

$$y_t + cy_s = Ky_{ss} \tag{8-7}$$

Provided we can treat c and K as constants this equation has the form of a classical diffusion equation and it is therefore amenable to analysis by classical methods. We shall show that the application of these methods can be extended to equation (8-6).

The assumptions made to derive (8-6) may be compared with those in the "kinematic wave" simplification of the Overland Flow equations. Lighthill and Whitham (1955) based their definitive work on kinematic waves on the assumption that discharge is a function of depth alone in a given cross-section, and as pointed out by Henderson (1966) p.367, this implies the assumption that the friction slope is equal to the fixed datum slope. That is, in our notation

$$S_f = z_s$$

Comparison of this equation with (8-1) and (2-57) indicates that the kinematic wave approach includes the same assumptions as the diffusion analogy approach, but also assumes $y_s \cos \phi$

to be small compared with $z_s$.

## 8.2 Diffusion Equation Solutions

The analytical solution which forms the basis of our discussion on the Diffusion Analogy was first presented by Hayami (1951). The relevant part of his paper is outlined and discussed in the notation of this thesis in Appendix F. However it is shown herein that a number of Hayami's assumptions, particularly that of a rectangular channel of constant width, unnecessarily restrict the applicability of his solution. Indeed we have already shown that (8-7) can be derived for any prismatic simple channel, and for laminar or turbulent flow, and thus, provided c and K are constant, the validity of Hayami's solution does not depend on the channel being rectangular or on the flow being turbulent. We now go further to show how Hayami's solution can also be applied to (8-6). This is important as most natural river channels gradually vary in width and datum slope, while lateral flows may also be significant.

We shall assume a solution to (8-6) of the form

$$y(s,t) = Y_0(s) + y_0(s) + f(s,t) \tag{8-8}$$

where the definitions of $Y_0$, $y_0$ and f will depend on the initial and boundary conditions specified. This assumed solution obviously owes its form to (F-3) used by Hayami. We reject any functional power series as it is impractical to assess more than the first term in such a series.

Substituting (8-8) into (8-6)

$$f_t + c(Y_0 + y_0 + f)_s + \frac{Kz_{ss}}{\cos\phi} + \frac{KS_f}{j\cos\phi}\frac{E_s}{E} - \frac{q}{B} = K(Y_0 + y_0 + f)_{ss} \quad (8\text{-}9)$$

We showed in Appendix F that in at least the case where the upstream boundary condition is two steady states separated by a sudden step change, (F-10) as well as (F-1) is satisfied by Hayami's solution (F-5). Now in (F-5) neither $\lambda$ nor the limits of the integral are dependent on s, so that if $y_0$ is a function of s alone it can be removed from under the integral sign as if it were constant. Thus (F-6), (F-7), (F-8), (F-9) and hence (F-10) are equally valid whether $y_0$ is a function of s or a constant. This indeed is the reason why the verification of Hayami's solution in Appendix F, including (F-18), was presented in relation to $(f_1/y_0)$. We accordingly rearrange (8-9) to the form

$$\left(\frac{f}{y_0}\right)_t + c\left[1 - \frac{2K(y_0)_s}{cy_0}\right]\left(\frac{f}{y_0}\right)_s - K\left(\frac{f}{y_0}\right)_{ss} = \frac{-F(s,t)}{y_0} \quad (8\text{-}10)$$

where by definition, from (8-9)

$$F(s,t) = c(Y_0)_s + c\left(1 + \frac{f}{y_0}\right)(y_0)_s + \frac{Kz_{ss}}{\cos\phi} + \frac{KS_f}{j\cos\phi}\frac{E_s}{E} - \frac{q}{B}$$
$$- K(Y_0)_{ss} - K\left(1 + \frac{f}{y_0}\right)(y_0)_{ss} \quad (8\text{-}11)$$

Now, from (F-10), we can clearly use Hayami's solution (F-6) if two conditions are met.

1. $Y_0$ and $y_0$ must be defined in such a way that $-F/y_0$ is approximately zero throughout the solution.

2. $\dfrac{2K(y_0)_s}{cy_0}$ must be small throughout the solution.

In order to find suitable definitions of $Y_0$ and $y_0$ we shall tentatively assume that (F-6) satisfies (8-10), using (F-17) as a convenient expression of (F-6). Thus we try

$$f = \frac{-2y_0}{\sqrt{\pi}} \int_0^H \exp\left[-(\frac{G}{2Z} - Z)^2\right] dZ \tag{8-12}$$

where G and H are again defined by (F-13) and (F-14), but $y_0$ is now a function of s rather than constant. Thus $f(s,0) = -y_0$ and $f(s,\infty) = 0$, and from (8-11)

$$F(s,0) = c(Y_0)_s + \frac{Kz_{ss}}{\cos\phi} + \frac{KS_f}{j\cos\phi}\frac{E_s}{E} - \frac{q}{B} - K(Y_0)_{ss} \tag{8-13}$$

$$F(s,\infty) = c(Y_0 + y_0)_s + \frac{Kz_{ss}}{\cos\phi} + \frac{KS_f}{j\cos\phi}\frac{E_s}{E} - \frac{q}{B} - K(Y_0 + y_0)_{ss} \tag{8-14}$$

Therefore, comparing (8-13) and (8-14) with (8-6), $F(s,0) = 0$ if $Y_0$ describes a steady state profile, and $F(s,\infty) = 0$ if $Y_0 + y_0$ also describes a steady state profile.

Thus it is clear from (8-13) and (8-14) that $Y_0$ is an initial steady state depth or stage at any point and $Y_0 + y_0$ is a final steady state stage at any point, corresponding respectively with the initial and final steady discharges introduced at the upstream boundary. Using (8-5) and (2-35), we write (8-11) as

$$\begin{aligned} F(s,t) &= (1 + \frac{f}{y_0})\left[c(y_0)_s - K(y_0)_{ss}\right] + c(Y_0)_s - K(Y_0)_{ss} \\ &\qquad + \frac{Kz_{ss}}{\cos\phi} + \frac{VAE_s}{BE} - \frac{q}{B} \\ &= (1 + \frac{f}{y_0})F_1 + F_2 \end{aligned}$$

say. Now $f(s,0) = -y_0$ and with $Y_0$ defined as above, $F(s,0) = F_2 = 0$. It is reasonable to assume that $VAE_s/BE$ and $q/B$ are no more dependent on time than are c and K, so that if c and K are taken as constant, as is necessary for the diffusion solution to hold, we can treat $F_2$ as independent of time. Alternatively if $VAE_s/BE$ and $q/B$ are not large, as is often true, $F_2$ is independent of time provided c and K are constant. In either case $F_2 = 0$ for all s,t. $F_1$ is also independent of time provided c and K are constant, so that, because $f(s,\infty) = 0$ and $F(s,\infty) = 0$ with $Y_0 + y_0$ defined as above, $F_1 + F_2 = 0$ also, and hence $F_1 = 0$ for all s,t. Thus $F(s,t) = 0$ for all s,t and our first condition for using Hayami's solution can be met.

Having established the meaning of $y_0$ we can turn to the second condition. Using (8-4) and (8-5)

$$\frac{2K(y_0)_s}{cy_0} = \frac{2j}{m+k}\,\frac{y\cos\phi}{S_f}\,\frac{(y_0)_s}{y_0} \tag{8-15}$$

We note from (8-1) and (2-28)

$$S_f = -h_s$$

Thus, using $\cos\phi = dx/ds$

$$\frac{S_f}{\cos\phi} = -h_x \tag{8-16}$$

Now for any steady state profile $y = Y(s)$, (8-6) becomes

$$\frac{m+k}{k}VY_s + \frac{jVY}{k}\frac{(S_f)_s}{S_f} + \frac{VY}{k}\frac{E_s}{E} - \frac{qY}{Ak} = 0$$

or

$$\frac{Y_s}{Y} = \frac{1}{m+k}\left[-\,j\frac{(S_f)_s}{S_f} - \frac{E_s}{E} + \frac{q}{AV}\right] \tag{8-17}$$

It is reasonable to suppose $(S_f)_s/S_f = h_{ss}/h_s$ is dependent only on s for all steady state profiles y = Y, and independent of Y. As $E_s/E$ is also independent of Y, as are m, k and j; and q/AV can generally be assumed reasonably constant or negligible, we are therefore able to assume that $Y_s/Y$ is approximately constant at a section, whence

$$\frac{(Y_0)_s}{Y_0} = \frac{Y_s}{Y} = \frac{(Y_0 + y_0)_s}{Y_0 + y_0}$$

Thus

$$\frac{(Y_0 + y_0)_s}{Y_0 + y_0} - \frac{(Y_0)_s}{Y_0} = 0$$

Whence

$$\frac{(y_0)_s}{y_0} = \frac{(Y_0)_s}{Y_0} = \frac{Y_s}{Y} \tag{8-18}$$

Using (8-18) and (8-16) in (8-15) we have

$$\frac{2K(y_0)_s}{cy_0} = \frac{-2jy}{(m+k)h_x}\,\frac{Y_s}{Y}$$

Now $Y_s$ is the rate of change of a steady state stage, while $h_x$ is any water surface slope which might occur at a section during the passage of the flood wave. Therefore $Y_s/h_x$ will be small except if the channel has a small datum slope (cf. (8-16) and (8-1)) and a large rate of change of width, and this case will be disregarded. Now $Y_s/h_x$ will be small for all Y, so we can choose Y = y at any section. Thus

$$\frac{2K(y_0)_s}{cy_0} = \frac{2jY_s}{(m+k)h_x} \leqslant \frac{2Y_s}{3h_x} \qquad (8\text{-}19)$$

because $k_{min} = 1$, so for laminar flow (j = 1, m = 2), $2j/(m+k_{min}) = \frac{2}{3}$, and for turbulent flow ($j = \frac{1}{2}$, $m = \frac{1}{2}$ to $\frac{2}{3}$), $2j/(m+k)_{min} = \frac{2}{3}$.

Hence our second condition is also satisfied, so that (8-12) will satisfy (8-10) and therefore (8-9).

Combining (8-8) and (8-12) we have

$$y(s,t) = Y_0(s) + y_0(s)\left\{1 - \frac{2}{\sqrt{\pi}}\int_0^H \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right]dZ\right. \qquad (8\text{-}20)$$

which is a solution of (8-6) subject to the given definitions of $Y_0$, $y_0$, H and G.

## 8.3 The Application of the Extended Solution

We have restricted our attention to the case of open channel flow when the flow at the upstream boundary takes the form of a series of steady states separated by sudden changes. We did this not only because the solution (F-6) is very much simpler in form than the more general (F-5), but also because such flows are of considerable practical interest as they commonly result from the operation of artificial controls such as the turbines in hydro-electric schemes. Now the function f behaves in such a way that the major part of the disturbance resulting from a sudden change in flow occupies a short length of the channel, with the diffusing leading and trailing edges behaving essentially as a series of steady states. Another impulse may thus be imposed on the solution after a reasonable time without appreciable mutual interference provided that it does not overtake the first impulse. This is assured if the same values of c and K hold for each of the impulses making up the flood wave, or if the step impulses are imposed a sufficient time apart for interaction between their effects to be negligible over the reach of the channel within which the solution is desired.

Hayami took this argument a stage further by superimposing the solutions resulting from a series of closely spaced step impulses which could be used to simulate any inflow to the upstream boundary, and this is reasonable if c and K behave as constants. Thus the flow resulting from any upstream boundary condition may be simulated by (8-20) provided $Y_0$, $y_0$, H and G can be determined in advance. In Hayami's discussion the determination of $Y_0$ and $y_0$ was not difficult because both were fixed at the upstream boundary. Often however, this assumption is not realistic as most channels vary gradually in width or shape, and this is why we have expressed $Y_0$ and $y_0$ as functions of s rather than constants.

If we recast (8-20) as

$$\frac{y(s,t) - Y_0(s)}{y_0(s)} = 1 - \frac{2}{\sqrt{\pi}}\int_0^H \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] dZ \qquad (8\text{-}21)$$

we see that we are interested in $Y_0$ only as a datum, so that any convenient value may be assigned to the initial (approximately) steady state depth. The final limiting change in depth $y_0$ must be obtained by finding the steady stage $Y_0(s) + y_0(s)$ required to pass through any section the steady discharge corresponding with the steady stage $Y_0(0) + y_0(0)$ at the upstream boundary. Equation (8-21) is therefore basically a description of the passage of a transient wave, marking the transition between two known steady states, down a channel. This of course also applies to Hayami's solution.

## 8.4 The Determination of Basic Diffusion Constants

Hayami did not discuss how c and K were to be evaluated, although he did suggest that K should be evaluated from experimental data rather than (F-2), particularly as J was so vaguely defined. Thomas and Wormleaton (1970) simply regarded c and K as lumped measures of the convective and diffusive characteristics of a river reach and found their values by a series of trial runs on a computer, selecting those values which were best able to reproduce the behaviour of an observed flood. We shall now show that c and K, and hence G and H, may be determined directly from an observed "calibration" flood by means of dimensionless graphs, provided this flood conforms approximately to the simple shape of a series of steady states separated by step changes. The principal application of this method is therefore in channels below artificial control structures, although subsequent flood hydrographs, to which the solution (8-21) is applied using the measured c and K, can take any shape at the upstream end of a calibrated channel reach.

Provided the step impulses are well spaced, there is no problem in finding $y_0$ for the "calibration" flood as the final steady stage can be measured directly, or at least the asymptote to which the stage is tending can be estimated, at all points on the channel.

Referring to (8-21) we let

$$p(s,t) = \frac{y - Y_0}{y_0} \tag{8-22}$$

where p can be regarded as the percentage at time $t_p$ of the final change in stage $y_0$ if $t_p$ is measured from the instant at which a change in stage was initiated as a step input at the upstream end of the channel.

Using (F-14), we may define

$$H_p = \frac{s}{2\sqrt{Kt_p}} \tag{8-23}$$

so that, from (8-21)

$$p = 1 - \frac{2}{\sqrt{\pi}}\int_0^{H_p} \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] dZ \tag{8-24}$$

which means that, given G and p we may compute $H_p$. Further, given $H_{p1}$, $H_{p2}$ corresponding to $p_1$, $p_2$ for any G, we can find, for any G, the ratio

$$\frac{t_{p1}}{t_{p2}} = \left[\frac{H_{p1}}{H_{p2}}\right]^2 \tag{8-25}$$

The dimensionless parameter G can be related to channel properties from (F-13), giving

$$G = \frac{cs}{2K}$$

$$= \frac{-(m+k)sh_x}{2jy} \qquad (8\text{-}26)$$

using (8-16). If c and K are constant, $-sh_x$ is a reference length related to the drop in elevation of the water surface between the upstream end of the channel and our observation point, and y is the depth. Reasonable values of G should therefore lie between $10^{-2}$ and $10^{4}$.

## 8.5 The Preparation of Dimensionless Graphs

$H_p$ and hence $t_{p1}/t_{p2}$ were computed for various values of G and p. G was varied over its probable useful range from $10^{-2}$ to $10^{4}$ by repeated multiplication by $10^{1/6}$, and p values of 5%, 25%, 50%, 75% and 95% were used. Trapezoidal Rule numerical integration was used to evaluate the integral, and each value of $H_p$ was estimated by linear interpolation immediately after the corresponding value of p was passed during the progressive evaluation of the integral. The step width dZ of the numerical integration was computed from specified changes, DFJ, in $\frac{G}{2Z} - Z$. Five values of DFJ (0.02, 0.01, 0.005, 0.005 and 0.01) were used successively, changing as the integration passed successive values of p. This was found to give ample accuracy.

The ratios of $t_p/t_{50}$ were calculated from the values of $H_p$ in each case using (8-25) and are tabulated against G and p in Table 8-1.

COMPUTATION OF CURVES TO EVALUATE DIFFUSION ANALOGY COEFFICIENTS C AND K FROM EXPERIMENTAL OBSERVATIONS.

M = 6, N = 37, AND DFJ = 0.020 0.010 0.005 0.005 0.010 SUCCESSIVELY

| G | P = 0.050 TP/TP(3) | 0.250 TP/TP(3) | 0.500 TP/TP(3) | 0.750 TP/TP(3) | 0.950 TP/TP(3) | 0.500 K FACTOR | 0.500 C FACTOR |
|---|---|---|---|---|---|---|---|
| 0.010 | 0.121 | 0.348 | 1.000 | 4.317 | 85.082 | 1.073948 | 0.021 |
| 0.015 | 0.122 | 0.350 | 1.000 | 4.246 | 75.639 | 1.062649 | 0.031 |
| 0.022 | 0.123 | 0.353 | 1.000 | 4.149 | 64.990 | 1.046485 | 0.045 |
| 0.032 | 0.125 | 0.357 | 1.000 | 4.018 | 53.806 | 1.023674 | 0.065 |
| 0.046 | 0.129 | 0.363 | 1.000 | 3.848 | 42.909 | 0.992006 | 0.092 |
| 0.066 | 0.133 | 0.372 | 1.000 | 3.636 | 33.073 | 0.949015 | 0.129 |
| 0.100 | 0.140 | 0.383 | 1.000 | 3.386 | 24.801 | 0.892461 | 0.178 |
| 0.147 | 0.149 | 0.399 | 1.000 | 3.108 | 18.267 | 0.820958 | 0.241 |
| 0.215 | 0.162 | 0.420 | 1.000 | 2.817 | 13.361 | 0.734996 | 0.317 |
| 0.316 | 0.180 | 0.447 | 1.000 | 2.532 | 9.810 | 0.637598 | 0.403 |
| 0.464 | 0.203 | 0.480 | 1.000 | 2.269 | 7.300 | 0.534330 | 0.496 |
| 0.681 | 0.234 | 0.519 | 1.000 | 2.037 | 5.548 | 0.432208 | 0.589 |
| 1.000 | 0.272 | 0.562 | 1.000 | 1.841 | 4.324 | 0.337923 | 0.676 |
| 1.468 | 0.318 | 0.607 | 1.000 | 1.679 | 3.465 | 0.256233 | 0.752 |
| 2.154 | 0.370 | 0.653 | 1.000 | 1.547 | 2.853 | 0.189284 | 0.816 |
| 3.162 | 0.427 | 0.698 | 1.000 | 1.442 | 2.412 | 0.136896 | 0.866 |
| 4.642 | 0.487 | 0.739 | 1.000 | 1.357 | 2.088 | 0.097387 | 0.904 |
| 6.813 | 0.546 | 0.777 | 1.000 | 1.289 | 1.848 | 0.068425 | 0.932 |
| 10.000 | 0.603 | 0.811 | 1.000 | 1.234 | 1.666 | 0.047639 | 0.953 |
| 14.678 | 0.656 | 0.840 | 1.000 | 1.191 | 1.528 | 0.032951 | 0.967 |
| 21.544 | 0.705 | 0.866 | 1.000 | 1.155 | 1.421 | 0.022685 | 0.977 |
| 31.623 | 0.748 | 0.888 | 1.000 | 1.127 | 1.338 | 0.015567 | 0.985 |
| 46.416 | 0.787 | 0.906 | 1.000 | 1.104 | 1.272 | 0.010659 | 0.989 |
| 68.129 | 0.820 | 0.922 | 1.000 | 1.085 | 1.220 | 0.007286 | 0.993 |
| 100.000 | 0.849 | 0.935 | 1.000 | 1.070 | 1.179 | 0.004976 | 0.995 |
| 146.779 | 0.873 | 0.946 | 1.000 | 1.057 | 1.146 | 0.003395 | 0.997 |
| 215.442 | 0.894 | 0.955 | 1.000 | 1.047 | 1.119 | 0.002316 | 0.998 |
| 316.226 | 0.912 | 0.963 | 1.000 | 1.039 | 1.097 | 0.001579 | 0.999 |
| 464.156 | 0.927 | 0.969 | 1.000 | 1.032 | 1.080 | 0.001076 | 0.999 |
| 681.288 | 0.939 | 0.975 | 1.000 | 1.026 | 1.065 | 0.000733 | 0.999 |
| 999.994 | 0.950 | 0.979 | 1.000 | 1.022 | 1.054 | 0.000500 | 1.000 |
| 1467.790 | 0.958 | 0.983 | 1.000 | 1.018 | 1.044 | 0.000341 | 1.000 |
| 2154.421 | 0.966 | 0.986 | 1.000 | 1.015 | 1.036 | 0.000232 | 1.000 |
| 3162.257 | 0.971 | 0.988 | 1.000 | 1.012 | 1.030 | 0.000158 | 1.000 |
| 4641.555 | 0.976 | 0.990 | 1.000 | 1.010 | 1.025 | 0.000108 | 1.000 |
| 6812.867 | 0.981 | 0.992 | 1.000 | 1.008 | 1.020 | 0.000073 | 1.000 |
| 9999.918 | 0.984 | 0.993 | 1.000 | 1.007 | 1.017 | 0.000050 | 1.000 |

Table 8-1

From (8-23)

$$K = \frac{1}{4H_{50}^{2}} \frac{s^2}{t_{50}} \tag{8-27}$$

Also, from (F-13) and (8-27)

$$c = \frac{G}{2H_{50}^{2}} \frac{s}{t_{50}} \tag{8-28}$$

$t_{50}$ is chosen as the standard $t_p$ because, being in the steepest region of the $p/t_p$ curve it is usually the best defined value of $t_p$, and in particular because c is commonly called the wave speed by analogy with kinematic wave theory (Section 8.1) and we might expect c to approximate to $s/t_{50}$. This proves to be the case as we see below.

Now for any value of G, $H_{50}$ is uniquely defined and hence the factor $1/4H_{50}^{2}$. This factor, which is the "K factor" listed in Table 8-1, will be denoted $F_K$.

Similarly, the "c factor" $F_c$ denotes $G/2H_{50}^{2}$.

and $$F_K = \frac{F_c}{2G} \tag{8-29}$$

We see from Table 8-1 that the c factor $F_c$ is always less than unity, but is asymptotic to unity as G tends to infinity. For practical purposes $F_c = 1$ for $G > 100$, justifying our expectations that c is closely related to the wave speed (see equation (8-28)).

The ratios of $t_p/t_{50}$ from Table 8-1 are plotted against G in Figure 8-1 for the most useful range of G, $0.1 \leqslant G \leqslant 1000$, while $F_c$ and $F_K$ are plotted against G for $0.01 \leqslant G \leqslant 100$ in Figure 8-2.

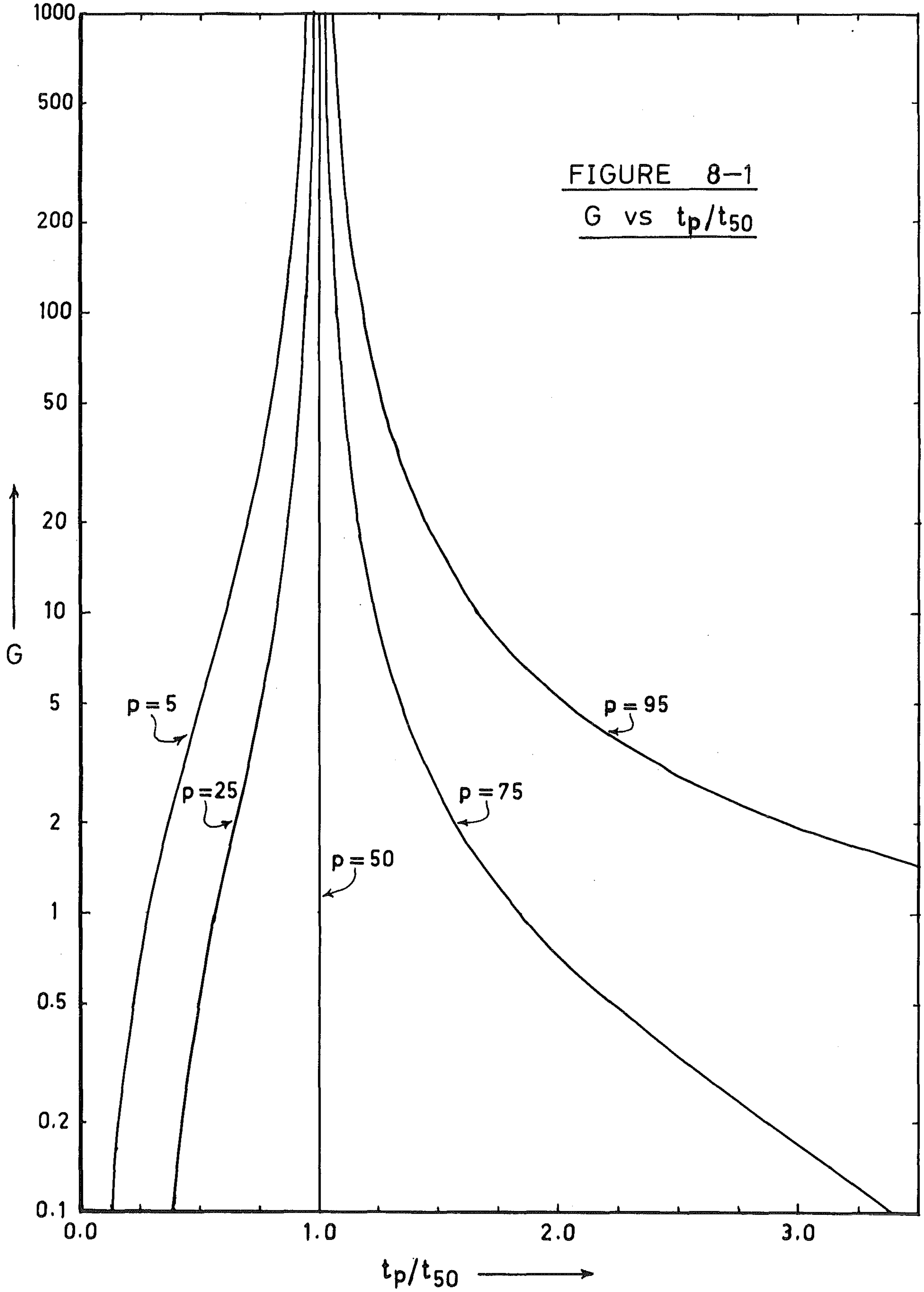


FIGURE 8-1
G vs $t_p/t_{50}$
1000
500
200
100
50
20
10
5
2
1
0.5
0.2
0.1
0.0
1.0
2.0
3.0
G
$t_p/t_{50}$
p = 5
p = 25
p = 50
p = 75
p = 95

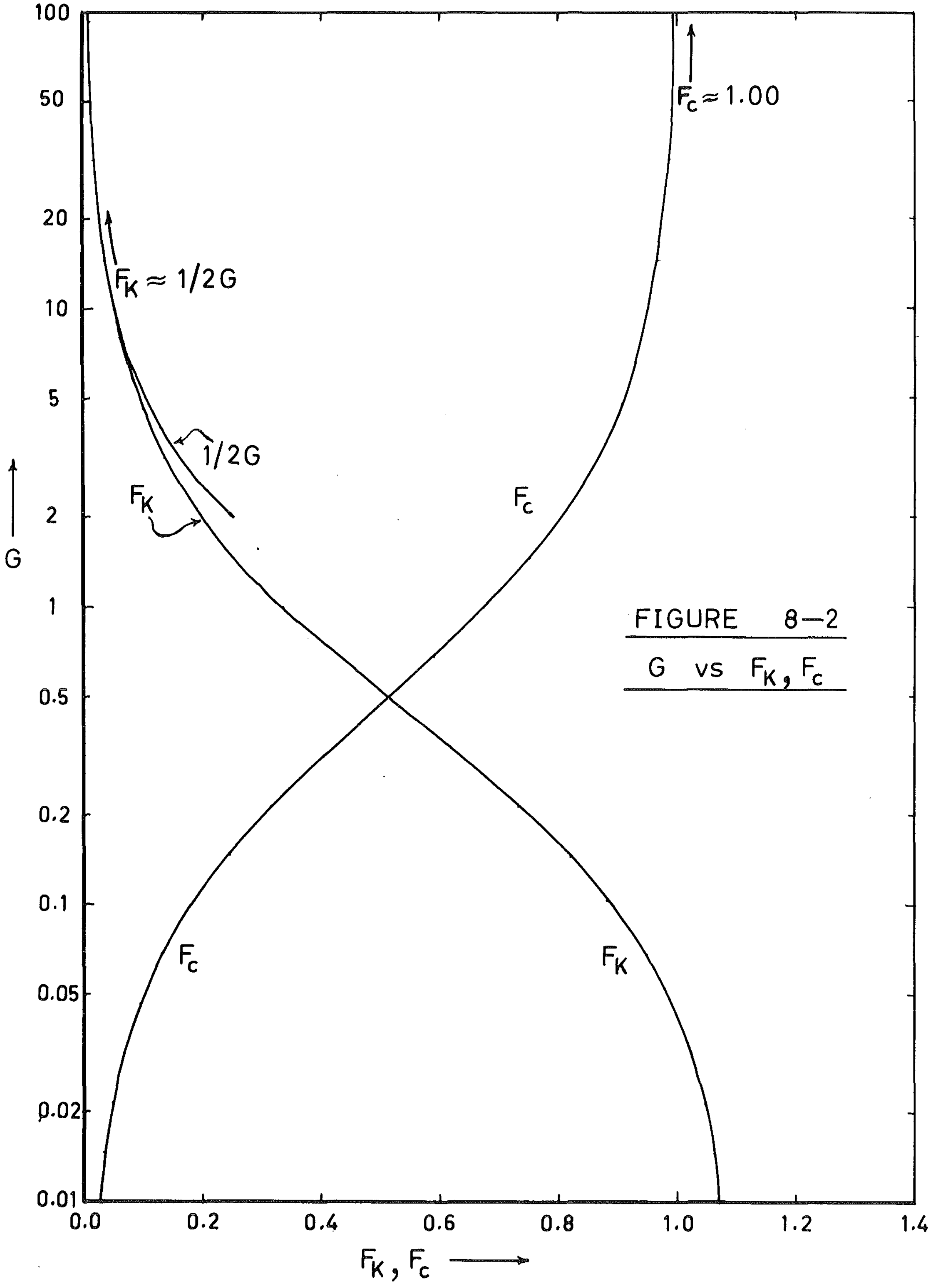



FIGURE 8—2

G vs $F_K$, $F_c$

For $G > 100$, $F_c = 1$, while $F_K$ can readily be estimated using (8-29). As indicated on Figure 8-2, $F_K$ is better estimated for $G > 10$ from $F_K \approx 1/2G$ than from Figure 8-2.

We may rewrite (8-27) as $G = \frac{F_c}{2K} \frac{s^2}{t_{50}}$ which indicates the effect of K in the description of the wave. Given an observation point at s and a value of $t_{50}$, we may still vary the wave profile by varying K. If K is increased, G will decrease, as $F_c$ will change in magnitude in the same sense as G. Smaller values of G correspond to flatter profiles (see Figure 8-1), so that the value of K is related to the steepness of the wave. That is, if K is small the wave is steep, while if K is large the wave is drawn out.

## 8.6 The Use of the Dimensionless Graphs

Now we may derive experimental values of c and K from observed results as follows. Experimental values of at least some of $t_5$, $t_{25}$, $t_{50}$, $t_{75}$ and $t_{95}$ can be obtained from the observed stage/time curve provided some estimate can be made of the final change in stage $y_0$. The ratios $t_5/t_{50}$, $t_{25}/t_{50}$, $t_{75}/t_{50}$ and $t_{95}/t_{50}$ can then be obtained. From these ratios a value of G may be obtained from Figure 8-1, as is demonstrated in Chapter 9. Given G, $F_c$ and $F_K$ can be obtained from Figure 8-2. Then, using

$$c = F_c \frac{s}{t_{50}} \tag{8-30}$$

$$K = F_K \frac{s^2}{t_{50}} \tag{8-31}$$

we obtain c and K using our known values of s and $t_{50}$.

The four experimental ratios $t_5/t_{50}$, $t_{25}/t_{50}$, $t_{75}/t_{50}$ and $t_{95}/t_{50}$ will rarely give a consistent value of G, so that some adjustment is normally needed. In the fitting procedure more weight will normally be given to the $t_{25}/t_{50}$ and $t_{75}/t_{50}$ values, but the effect of departures at the upstream boundary from the ideal step impulse should also be considered. The recommended fitting procedure is best described with examples and hence is left to Chapter 9.

### 8.7 Important Diffusion Parameters

So far we have regarded c and K as the required channel parameters, but it is now apparent that the relationship between p and $t_p$ is of more direct interest as these two alone are sufficient to describe the stage/time curve between the two known steady stages $Y_0$ and $Y_0 + y_0$ at any observation point. If c and K are constant for a range of stages in a channel, then G will be constant at a given observation point. Hence $H_p$ and therefore $t_p$ will be constant for any p at that point. Thus any transition within the allowable range of stage will, according to our assumptions, change a percentage p of its final change in stage at constant time $t_p$ from the instant that the transition was initiated as an impulse at the upstream boundary.

Now if we can measure sufficient values of $t_p$ at an observation point to fit a value of G and hence find a best fit value of $t_{50}$ as described in Section 8.6, we can draw a $p/t_p$

graph for that point by using Figure 8-1. This means that provided some calibration step impulse causes a wave in a channel to which a reasonably consistent value of G can be fitted for each observation point, the diffusion solution enables us to predict that the $p/t_p$ curves obtained at each observation point can be applied to the effects of all other upstream impulses, at least in the same range of stage. This greatly simplifies computation even if the effects of several impulses are to be superimposed as suggested by Hayami.

# CHAPTER 9

## EXPERIMENTAL VERIFICATION OF DIFFUSION THEORY

### 9.1 Introduction

In this chapter we apply the Diffusion theory of Chapter 8 to the records of two floods in the Clutha River, New Zealand. The floods were produced artificially by the operation of the Roxburgh hydro-electric power station, and were recorded by parties of final year Civil Engineering undergraduates from the University of Canterbury. The first flood of 5th October 1966 is used as a calibration flood, and the travel times $t_p$ for p = 5%, 25%, 50%, 75% and 95% are obtained. These travel times are applied to the subsequent flood of 26th August 1968 in order to predict stage/time curves at observation points along the channel, and good agreement is found between the predicted and observed curves for that flood. Because we are interested principally in testing our Diffusion theory, we do not attempt to follow the waves further downstream than the Millers Flat bridge, as at this point the negative wave is just beginning to interfere with the positive wave. This means we are able to use experimental values of $y_0$ which eliminates any errors arising from the estimation of $y_0$ and hence should isolate any errors in the Diffusion Analogy theory itself.

The three observation points were:

1. Roxburgh Dam, using the hydro station's tailwater level recorder.
2. Roxburgh Bridge, 5.5 river miles downstream of the dam.
3. Millers Flat Bridge, 16.4 river miles downstream of the dam.

## 9.2 The Calibration Flood Wave

The three stage/time records of the 1966 flood are plotted to the same scale in Figure 9-1, with the initial steady state $Y_0$ as datum. This $Y_0$ corresponds with a flow of 1600-1700 cusecs which had been maintained for almost seven hours since 23.00 hours on the day before.

We assume that each step change in stage occurs at the instant when one half of the final change in stage $y_0$ has occurred. Figure 9-1, curve (a) therefore gives an initiation time of 7.10 hours for the positive wave, and 10.10 hours for the negative wave, where the time scale is in decimal hours from midnight. The required percentages of the final changes in stage are plotted in Figure 9-1 on curves (b) and (c). In order to determine $y_0$ for the negative wave, it is necessary to extrapolate the stage/time curves to the final steady state stages. This extrapolation is shown dotted. The time corresponding to each percentage has been read off and is marked in Figure 9-1, below curves (b) and (c). The initiation time of each wave is subtracted from these times to give the experimental travel times in Table 9-1, e.g. 7.17 minus 7.10 gives 0.07 hours.

FIGURE 9–1

The Calibration Flood Wave

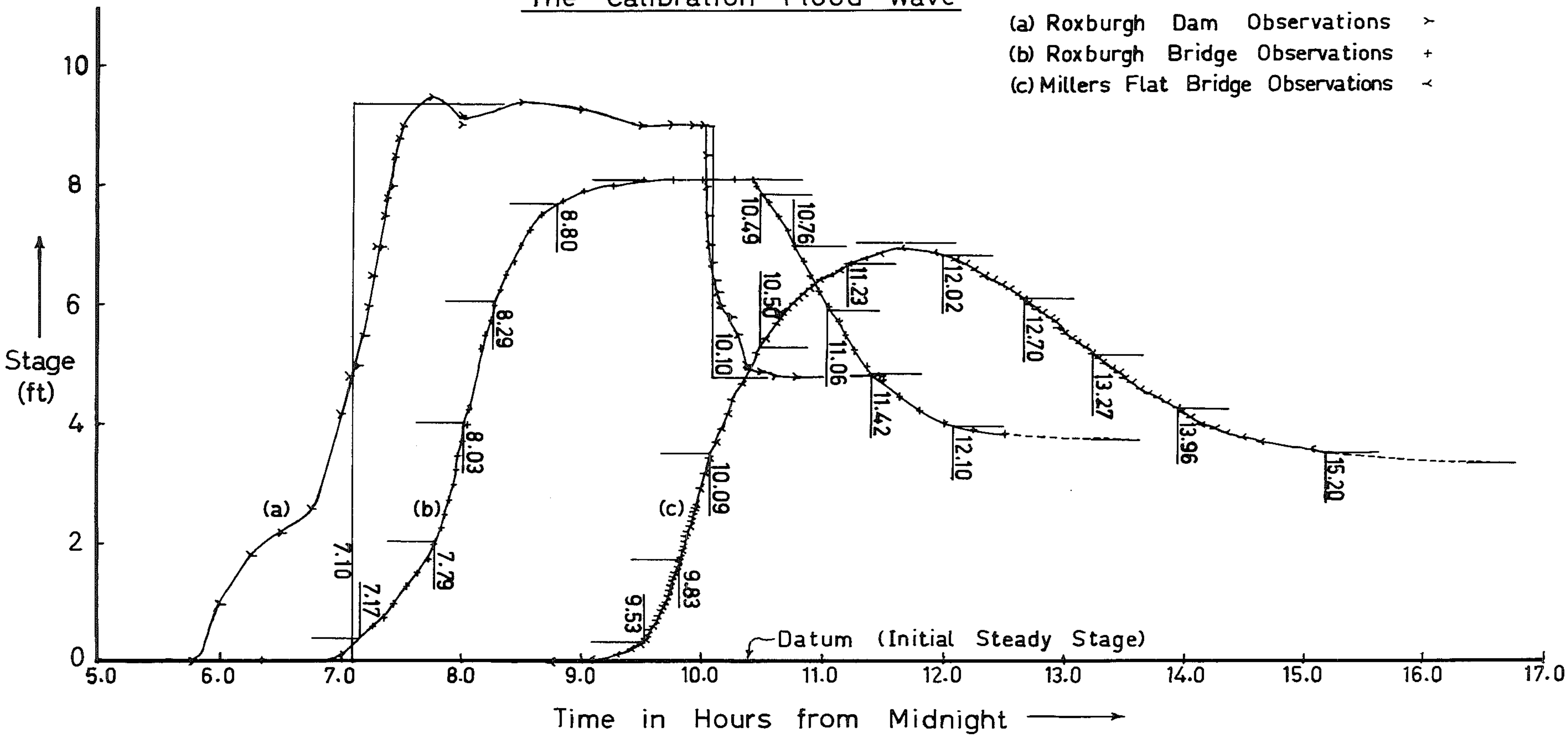


Table 9-1

The Determination of Best Fit Values of G and Travel Times

| | Roxburgh Bridge | | | | Millers Flat Bridge | | | |
|---|---|---|---|---|---|---|---|---|
| | Positive Wave | | Negative Wave | | Positive Wave | | Negative Wave | |
| | Expt | Fitted | Expt | Fitted | Expt | Fitted | Expt | Fitted |
| $t_5$ | 0.07 | 0.50 | 0.39 | 0.42 | 2.43 | 2.29 | 1.92 | 1.93 |
| $t_{25}$ | 0.69 | 0.71 | 0.66 | 0.66 | 2.73 | 2.71 | 2.60 | 2.60 |
| $t_{50}$ | 0.93 | 0.92(3) | 0.96 | 0.93 | 2.99 | 3.01 | 3.17 | 3.17 |
| $t_{75}$ | 1.19 | 1.19 | 1.32 | 1.30 | 3.40 | 3.37 | 3.86 | 3.86 |
| $t_{95}$ | 1.70 | 1.72 | 2.00 | 2.07 | 4.13 | 3.97 | 5.10 | 5.17 |
| $t_5/t_{50}$ | 0.08 | 0.54 | 0.41 | 0.45 | 0.81 | 0.76 | 0.61 | 0.61 |
| $t_{25}/t_{50}$ | 0.74 | 0.77 | 0.69 | 0.71 | 0.91 | 0.90 | 0.82 | 0.82 |
| $t_{75}/t_{50}$ | 1.28 | 1.29 | 1.37 | 1.40 | 1.14 | 1.12 | 1.22 | 1.22 |
| $t_{95}/t_{50}$ | 1.83 | 1.87 | 2.08 | 2.23 | 1.38 | 1.32 | 1.61 | 1.63 |
| $G_{t5}$ | Note(1) | 6.5 | 2.8 | 3.8 | 62 | 35 | 10.0 | 10.5 |
| $G_{t25}$ | 4.8 | 6.5 | 2.9 | 3.8 | 48 | 35 | 10.5 | 10.5 |
| $G_{t75}$ | 6.8 | 6.5(2) | 4.3 | 3.8 | 28 | 35 | 11.0 | 10.5 |
| $G_{t95}$ | 7.0 | 6.5 | 4.7 | 3.8 | 25 | 35 | 11.5 | 10.5 |

Notes: (1) $G_{t5}$ neglected as obviously affected by the departure of the input wave from the ideal step wave.

(2) More weight placed on $G_{t75}$ and $G_{t95}$ as these least affected by the departure of the input wave from the step wave.

(3) $t_{50}$ adjusted to give least squares departure of the fitted $t_p$ values from the more reliable experimental values of $t_p$.

In Table 9-1, the experimental travel times are formed into the ratios $t_5/t_{50}$, $t_{25}/t_{50}$, $t_{75}/t_{50}$ and $t_{95}/t_{50}$ and the value of G corresponding with each ratio is read off Figure 8-1. A weighted average G is then selected and the corresponding fitted ratios $t_5/t_{50}$, $t_{25}/t_{50}$, $t_{75}/t_{50}$ and $t_{95}/t_{50}$ are read off Figure 8-1. If most of these ratios are larger (smaller) than the experimental ratios as in the first two columns, then $t_{50}$ must be reduced (increased) slightly. If we do not adjust $t_{50}$ most of the values of $t_p$ will be too high (low), whereas we wish to fit a curve such that the experimental values of $t_p$ lie equally on either side. It is not difficult to assess the adjustment needed to $t_{50}$ in order to produce a least squares fit through the three most reliable values of $t_p$, with reasonable weight being attached to the other two values.

Large changes in G produce relatively small change in the ratios of the travel times so that a relatively large error in G will have little effect on the fitted values of $t_p$. For instance, if we select 40 instead of 35 (a 14% error) for the fitted average G value for the positive wave at Millers Flat, the fitted travel times become respectively 2.32, 2.71, 3.01, 3.37 and 3.88 hours, on average about 1% different from the travel times corresponding with a G value of 35. It can be seen that G = 35 gives a slightly better correspondence with the experimental travel times than does G = 40.

Under our assumptions, c and K should not vary between a positive and negative wave and hence G should have a single value for an observation station. However, from (8-1) $S_f$ may vary markedly with the sign of $y_s$, so that it is possible to derive different values of G for positive and negative waves and to state that G is approximately constant for a given wave type at an observation point without inconsistency, provided the waves are appreciable in height.

Table 9-1 shows that we can fit Hayami type solutions to the experimental stage/time records from the 1966 Clutha River flood, with the average discrepancy between the observed and fitted travel times of the order of 2%. The values of G appear to be determinable with an error of less than $\pm 10\%$.

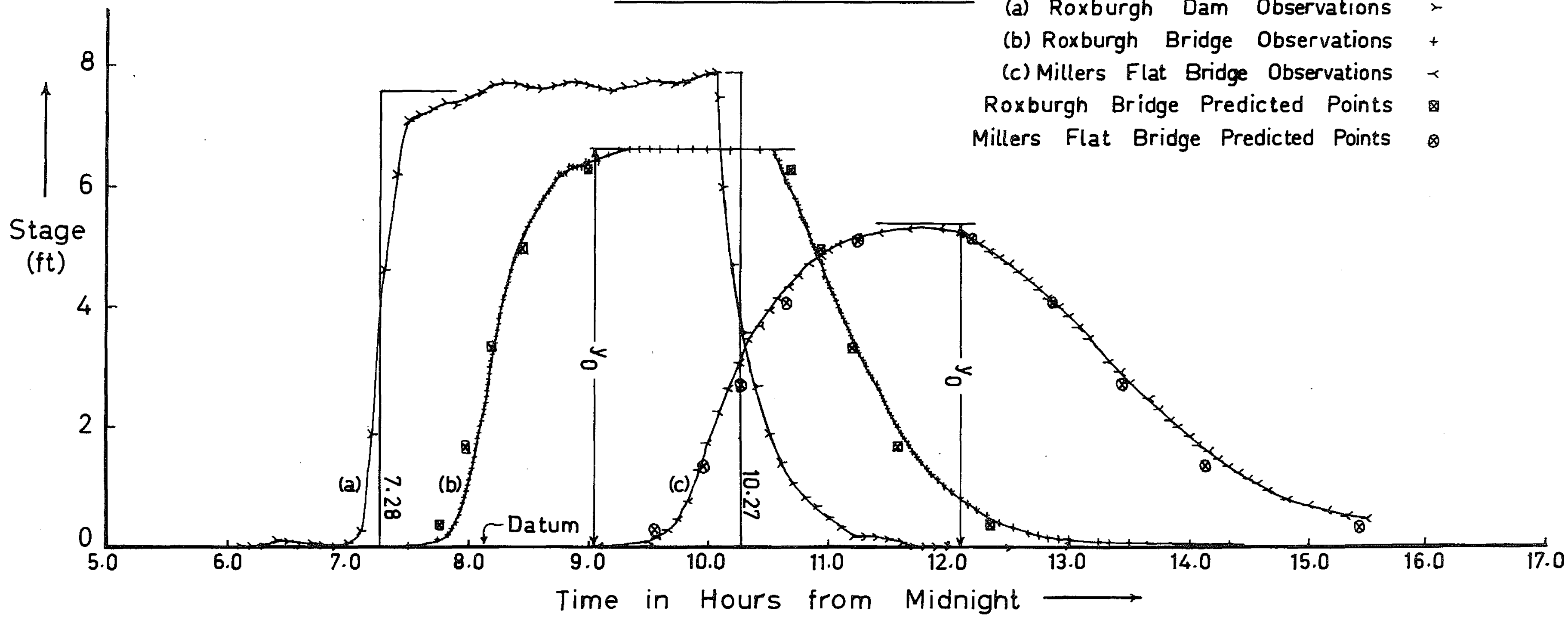
## 9.3 The Test Flood Wave

The stage/time records of the 1968 flood are plotted in Figure 9-2. This flood differs from the 1966 flood in three important ways:

1. The initial flow on which this flood is imposed is 3500 cusecs as against 1650 cusecs for the 1966 flood.
2. The stage at the dam is increased essentially in one step over about $\frac{1}{2}$ hour as against the two steps over $1\frac{3}{4}$ hours for the 1966 flood.
3. The final steady state discharge after the abrupt decrease in flow is again 3500 cusecs, compared with 6,700 cusecs for the 1966 flood.

FIGURE 9—2

The Test Flood Wave



Therefore, although both maximum discharges are about 17,500 cusecs, the two floods are appreciably different and hence the prediction of the behaviour of the 1968 flood from that of the 1966 flood is by no means a trivial exercise.

Once again we assume that the time at which each wave originates is the instant when one half of $y_0$ has occurred at the dam, giving 7.28 hours for the positive wave and 10.27 hours for the negative wave. We then simply add the fitted travel times from Table 9-1 to give the abscissae of the predicted points (marked x in Figure 9-2) with the ordinates being supplied by the appropriate percentages of the experimental values of $y_0$ at each observation point. Good agreement is obtained (Figure 9-2) between the stage/time curve predicted by joining the crosses and the observed stage/time curve at both Roxburgh Bridge and Millers Flat Bridge. It is apparent from the Millers Flat hydrograph that the flow returns to steady depth more slowly than predicted, which points to a small error in the extrapolation of the 1966 Millers Flat hydrograph, and hence in the final steady stage deduced from that extrapolation. If we assumed a final stage 0.1 ft lower than previously for the 1966 flood, we would increase the values of $t_p$ slightly for the negative wave to Millers Flat, in particular $t_{75}$ and $t_{95}$. This would make the fit of the predicted negative wave at Millers Flat even closer. In other words continuation of the observations at Miller's Flat to the steady state would probably have improved our predictions slightly.

Note that by using an experimental value of $y_0$ we did not require any channel properties other than $s$.

## 9.4 Comparison of Fitted and Derived c and K Values

We wish to compare the values of c and K corresponding with the fitted curves (say $c^1$ and $K^1$) with those derived from (8-4) and (8-5) which we shall continue to denote c and K. $c^1$ and $K^1$ are therefore derived using Figure 8-2 and equations (8-30) and (8-31). The results are given in Table 9-2.

For $G > 3$, Figure 8-2 indicates errors of less than $1\frac{1}{2}\%$ for $F_c$ and about 10% for $F_K$ correspond with 10% error in G. Therefore, assuming maximum errors of $\pm 10\%$ for G, $\pm 3\%$ for $t_{50}$ and $\pm 1\%$ for s, we arrive at a maximum error of about $\pm 5\%$ for c derived from (8-30)(i.e. $c^1$) and $\pm 15\%$ for K derived from (8-31) (i.e. $K^1$). These errors are considerably smaller than those to be expected from the use of (8-4) and (8-5) for the evaluation of c and K.

To use (8-4) and (8-5) we need data on channel properties. Values derived from cross sections measured at the three observation points are shown in Table 9-3. $y_{min}$ is the maximum depth at the low steady state, $y_{max}$ is the maximum depth at the high steady state, k is the average channel exponent as defined in equation (2-37), $V_{min}$ is the velocity at the low steady state, and $V_{max}$ is the velocity at the high steady state. k was obtained from Figure 9-3, which is a plot of the area of each cross section up to at least three different horizontal lines, chosen to

<u>Table 9-2</u>

The Determination of $c^1$ and $K^1$

| | Roxburgh Bridge | | Millers Flat Bridge | |
|---|---|---|---|---|
| | Positive Wave | Negative Wave | Positive Wave | Negative Wave |
| Distance s (miles) | 5.5 | 5.5 | 16.4 | 16.4 |
| $t_{50}$ (hours) | 0.92 | 0.93 | 3.01 | 3.17 |
| G | 6.5 | 3.8 | 35 | 10.5 |
| $F_c$* | 0.93 | 0.89 | 0.99 | 0.95 |
| $s/t_{50}$ | 5.98 | 5.92 | 5.45 | 5.17 |
| $c^1$(miles/hour) | 5.6 | 5.3 | 5.4 | 4.9 |
| $F_K$* | 0.075 | 0.115 | 0.0143 | 0.045 |
| $s^2/t_{50}$ | 32.9 | 32.6 | 89.3 | 84.8 |
| $K^1$(miles$^2$/hr) | 2.5 | 3.8 | 1.3 | 3.8 |

* Read off Figure 8-2

encompass the actual range of water surface levels, against the maximum depth y of the bed below each line. Each area was obtained by graphical integration (i.e. counting squares) on an undistorted plot of each cross section on a scale of 1 inch to 50 feet. The cross sectional area corresponding with each $y_{min}$ and $y_{max}$ were read off the best fit lines in Figure 9-3 and were divided into the known steady state discharges, 3,500 cusecs and 17,500 cusecs respectively, to give $V_{min}$ and $V_{max}$ at each observation point.

R.L. (Invert) is the level of the lowest point in the cross section relative to Mean Sea Level, whence $z_s$ follows as we know the distance s of the observation stations from the dam. Cos $\phi$ may of course be taken to be unity. R.L.($Y_0$) is the reduced level of the water surface for the low steady state at the observation point, and $-h_{0s}$ is the corresponding water surface slope, which from (2-28) can also be written $z_s - Y_{0s}$.

The differences between $z_s$ and $z_s$-$Y_{0s}$ bring out the point discussed in Chapter 2 that when we wish to evaluate the pressure forces acting on a flow, the slope $z_s$ of the datum from which y is measured is of little concern, as it is the water surface slope in which we are interested. Thus we use the initial steady state stage slope $z_s$-$Y_{0s}$ as the fixed reference slope represented by $z_s$ in equation (8-1) so that "$y_s$" will now refer to the slope of the water surface relative to $z_s$-$Y_{0s}$. This does not in fact make any difference if (8-1) is used to evaluate $S_f$, as the water

Table 9-3

Flow Properties and Bed Slopes

| | Roxburgh Bridge | Slope $x10^4$ | Roxburgh Dam | Slope $x10^4$ | Millers Flat Bridge |
|---|---|---|---|---|---|
| $y_{min}$* | 10.0 ft | | 22.8 ft | | 5.8 ft |
| $y_{max}$* | 16.6 ft | | 30.6 ft | | 11.1 ft |
| k* | 2.0 | | 2.2 | | 1.5 |
| $V_{min}$* | 4.4ft/sec | | 2.0ft/sec | | 3.3 ft/sec |
| $V_{max}$* | 8.3ft/sec | | 5.3ft/sec | | 6.0 ft/sec |
| R.L.(Invert)* | 244.9 ft | | 252.2 ft | | 198.4 ft |
| $z_s$* | | 2.5 | | 6.2 | |
| R.L.($Y_0$)* | 254.9 ft | | 275.0 ft | | 204.2 ft |
| $-h_{0s} = z_s - Y_{0s}$ | | 6.9 | | 8.2 | |

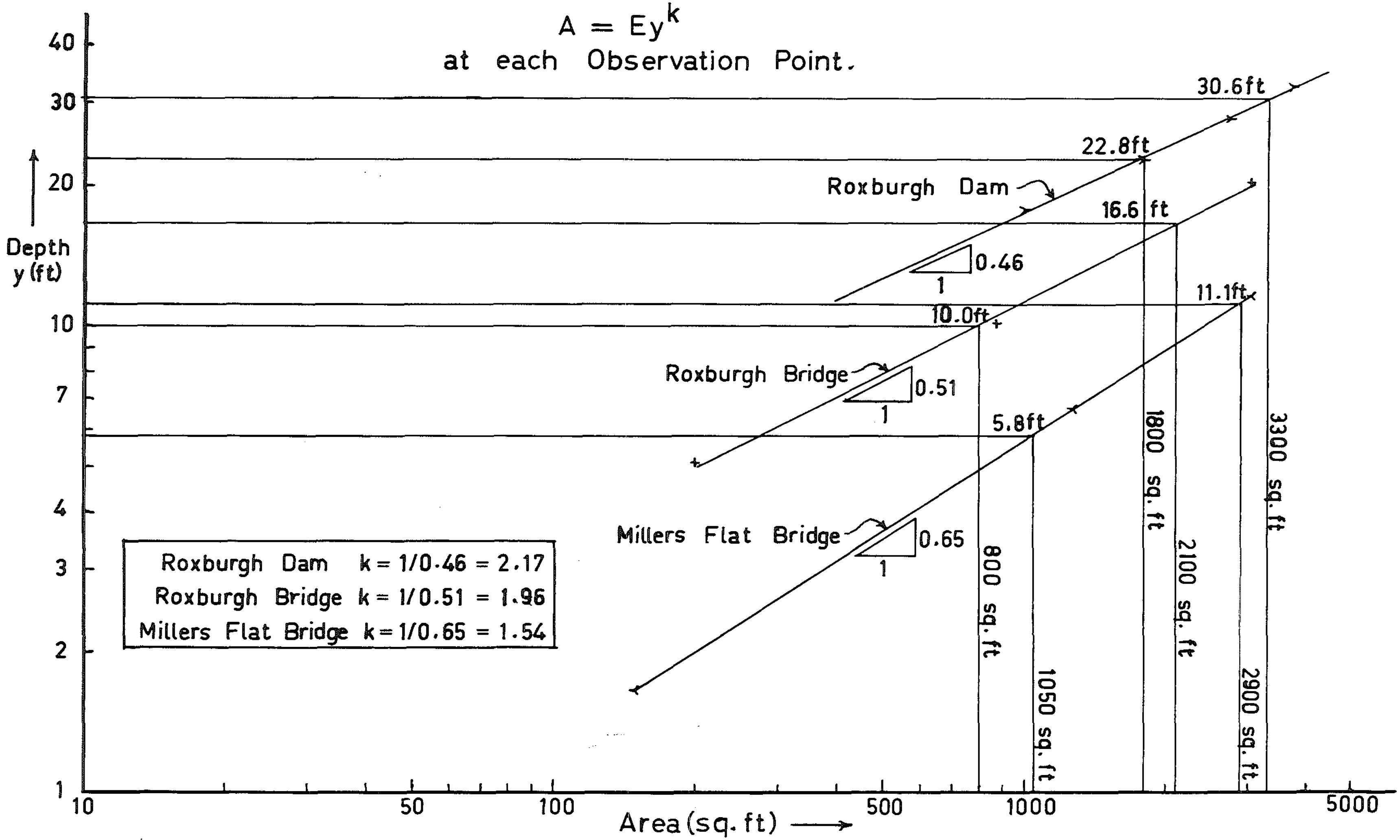
* See text for explanation

FIGURE 9—3

The Area/Depth Relationship

$A = Ey^k$

at each Observation Point.



surface slope is independent of the fixed slope to which $y_s$ is referred, but if $S_f = z_s$ is used as in the kinematic wave simplification (see Section 8.1), it is clearly important to read $z_s$ as meaning some typical water surface slope.

The parameters in Table 9-3 are used to derive c and k in Table 9-4. In the reach to Roxburgh Bridge more weight is placed on the k, V, y from the Roxburgh Bridge cross section as being more typical of the reach than the Roxburgh Dam cross section, and similarly in the reach from the dam to Millers Flat Bridge more weight is given to the k, V and y from the Millers Flat Bridge cross section. A slightly larger average velocity is assumed for the positive waves than for the negative waves as follows from equations (2-54) and (8-1). The average $y_t$ for each wave at each observation point is taken as the gradient of the straight line joining the $t_5$ and $t_{95}$ points on each stage/time curve, from which weighted lumped averages of $y_t$ are assessed. $y_s$ is then obtained from $cy_s = -y_t$ which follows if we assume that the wave as a whole travels at a speed of c, and we use $y_s$ to find $S_f$ from

$$S_f = z_s - Y_{0s} - y_s$$

which corresponds with equation (8-1) as discussed above.

Clearly the values of k, V, y and $S_f$ cannot be averaged over a reach without introducing the possibility of large errors. Further, the best estimate of an average $y_s$ from our data involves the assumption that the wave as a whole travels at approximately the velocity c, which is not consistent with our diffusion analogy

Table 9-4

The Derivation of c and K from Channel Properties

| | To Roxburgh Bridge | | To Millers Flat Bridge | |
|---|---|---|---|---|
| | Positive Wave | Negative Wave | Positive Wave | Negative Wave |
| Average k* | 2.0 | 2.0 | 1.7 | 1.7 |
| Average V* (miles/hour) | 4.0 | 3.8 | 3.5 | 3.3 |
| $c = (1 + \frac{2}{3k})V$ (miles/hour) | 5.3 | 5.1 | 4.9 | 4.6 |
| Average $y_t$* (ft/hour) | 7.0 | -4.0 | 4.0 | -2.0 |
| $y_s \approx -y_t/c\mathrm{x}5280$* | $-2.5\mathrm{x}10^{-4}$ | $1.5\mathrm{x}10^{-4}$ | $-1.5\mathrm{x}10^{-4}$ | $0.8\mathrm{x}10^{-4}$ |
| Average $S_f$* | $9.4\mathrm{x}10^{-4}$ | $5.4\mathrm{x}10^{-4}$ | $9.7\mathrm{x}10^{-4}$ | $7.4\mathrm{x}10^{-4}$ |
| Average y* (miles) | $2.5\mathrm{x}10^{-3}$ | $2.5\mathrm{x}10^{-3}$ | $1.9\mathrm{x}10^{-3}$ | $1.9\mathrm{x}10^{-3}$ |
| $K = \frac{Vy}{2kS_f}$ (miles$^2$/hour) | 2.7 | 4.4 | 2.0 | 2.5 |

* See text for explanation

unless $y_{ss}$ is small. Therefore, all the quantities, in particular c and K, in Table 9-4 must be regarded as crude estimates. It is clear however that the $c^1$ and $K^1$ obtained by our new fitting procedure conform in magnitude to the definitions of c and K in equations (8-4) and (8-5) as it is possible to find reasonable average values of k, V, y and $S_f$ which will produce $c^1$ and $K^1$ from those equations.

It is noticeable that the y and V necessary to match K to $K^1$ at the Millers Flat Bridge must be greater for the negative than for the positive wave. Of course the average depth and velocity are higher for the negative wave than the positive wave for the 1966 calibration flood, but if anything K seems slightly larger for the negative wave of the 1968 flood than for that of the 1966 flood, and in the 1968 flood the average y and V should be similar for both waves. The true explanation lies in the fact that the diffusion of the wave does not depend only on the derivatives of y, as is suggested by our assumption that K is a constant, but also on y itself because the flow velocity increases with depth. This factor tends to steepen positive waves and flatten negative ones. As already discussed in Section 8.5 a steeper wave reflects a smaller K, so that this explanation is consistent with the observed behaviour of the waves.

## 9.5 Discussion of the Diffusion Theory Assumptions

We return to the assumption in Section 8.1 that the omission of the acceleration and extra momentum terms from (2-57) is acceptable. The Diffusion Analogy solution gives good agreement with experimental results reported herein, but it is still of interest to assess the magnitude of the omitted terms, at least in the floods under investigation. Both these floods were caused by the operation of artificial controls and as appreciable lateral inflow might be expected to the reach under study only as the result of rain, q, and hence the extra momentum term qU/A, can be neglected. For turbulent flow $\beta$ and $\eta$ are both nearly unity, so we wish to evaluate the terms

$$VV_s + V_t = \frac{dV}{dt} \text{ if we "follow the fluid"}$$

Because the wave overtakes the fluid at a slower rate than it passes a fixed observation point, we expect $V_t$ to be greater in magnitude than dV/dt. This is not necessarily true if the channel is subject to rapid changes in width or $V_t$ is small, but neither of these difficulties are apparent in the floods under consideration. Now by inspection of the stage/time records of the floods, the maximum values of $V_t$ will occur in the positive wave of the 1968 flood near the dam. Referring to Table 9-3 and Figure 9-2, an average $V_t$ at the dam would be

$$\frac{5.3 - 2.0 \text{ ft/sec}}{0.4 \text{ hrs}} = 2.3 \times 10^{-3} \text{ ft/sec}^2$$

and at the Roxburgh Bridge

$$\frac{8.3 - 4.4 \text{ ft/sec}}{0.7 \text{ hrs}} = 1.5 \times 10^{-3} \text{ ft/sec}^2$$

or less than 10% of the magnitude of $gS_f \approx 30 \times 10^{-3}$ ft/sec$^2$.

Now it is true that some local $V_t$, particularly near the leading edge of the positive wave, may be considerably greater than this "average", which was derived using steady state velocities. However, dV/dt may still be small and also any $V_t$ greatly exceeding the average must apply to a very restricted time period. This matter of the relative magnitude of the acceleration and slope terms is further discussed by Henderson (1966) Chapter 9.

## 9.6 Applications of Diffusion Theory

As shown in Section 9.3 the diffusion theory developed in Chapter 8 can lead to quite accurate predictions of the behaviour of waves produced by the operation of artificial controls. Such behaviour should be equally predictable by the numerical methods discussed in Part II, but for these to be applicable to flows in natural channels a great deal of effort must be expended on defining all properties of the channel, including the bed geometry, roughness, the effects of bends, expansions, bridges and so on. In contrast a steady state stage/discharge rating curve at each cross section at which the wave behaviour is required is the only information required by these diffusion methods. If the upstream discharge input can be measured, as at any artificial control, even these rating curves can be assessed by direct observation, at least in the normal range of flows. Further, these methods do not rely on large computing facilities.

The use of factors measured at normal flows to predict flood behaviour at extreme flows is of course inaccurate. However from the results of Section 9.3 we are equally as entitled to regard $p/t_p$ curves as channel properties as we are Manning's n for instance. Hence it is quite possible that such inaccuracies would affect other methods as much as diffusion methods.

Unfortunately, the increase in flow velocity with increase in depth tends to make a closely spaced series of waves interact in a more complex way than the simple superposition suggested by Hayami. Thus, initially at least, diffusion methods are probably best restricted to routing floods in channels below artificial control structures through which discharge tends to vary in a series of finite increments. The interaction between the fringes of the resulting series of waves may then be represented adequately by superposition.

The diffusion methods discussed in Part III are therefore unlikely to contribute greatly to the prediction of the actual peak stage corresponding to a known reasonably steady upstream discharge. Such a peak stage is best read off the appropriate stage/discharge curve. Rather these methods provide a simple, rapid means of predicting the rate of rise of a flood at an observation point from the discharge/time relationship at the upstream control, and should therefore enable appropriate emergency action to be taken to minimise the damage resulting from an extreme flood.

# PART IV

# CHAPTER 10

# SUMMARY AND CONCLUSIONS

## 10.1 Summary

Overland flow equations are introduced in Chapter 1 as a system of equations resulting from the treatment of surface water flows by a one dimensional analysis using the continuity principle and Newton's second law of motion. In Chapter 2 such analysis is applied, involving assumptions that all streamlines in the flow are approximately straight and parallel and that the bed resistance to flow may be described either by simple laminar flow concepts or by semi-empirical relationships like the Manning formula. Common assumptions discarded from the more usual overland flow analysis include the assumptions that the flow is near horizontal, that the velocity is uniform over any cross section, that the lateral inflow is small compared with the channel flow, that the channel is prismatic, and that a "bed slope" is essential as a reference slope. It is also shown that even though the surface over which the flow passes is assumed fixed, provision could be made in the analysis for bed movement if some suitable relationship was made available between the water surface slope and mean velocity and section area.

The above analysis produces a more general form of the St Venant equations, called in this thesis the Overland Flow equations. This system of equations is then reduced to characteristic form in Chapter 3. As maximum generality of the equations is maintained to this point, a general form of wave celerity c and "stage variable" w appear and their functions are clarified. The geometry of a prismatic simple channel, defined in Chapter 2, is shown to permit the expression of the characteristic form of the Overland Flow equations explicitly as two equations in two dependent variables, mean velocity V and celerity c.

Methods of solution of wave propagation systems are then discussed, and the work of Lax, Courant, Isaacson and Rees is outlined. It is pointed out that the proof given by Courant, Isaacson and Rees (1952) of boundedness of the error between a difference solution and the exact solution applies only for small time increments, which suggests an explanation for the disappointing practical results which have led to the virtual abandonment of their simple difference scheme.

Consistency, convergence and stability are then introduced, following Richtmyer (1957), where it is shown that together consistency and stability imply convergence. The Fourier series method of stability analysis advocated by Richtmyer (1957) is outlined, and is used in Appendix A to produce the same "Courant condition" for stability as the analysis of Courant, Isaacson and Rees when applied to their difference scheme. Again, unfortunately, the analysis applies only for small time steps, and again the implied limit on the time step cannot be evaluated quantitatively. Little attention appears to have been paid to the range of application of these analyses, so that a "theoretically convergent" difference solution may be found to be unobtainable in practice because the mesh refinement required is impracticable. At the beginning of Chapter 4 examples are quoted from the work of Stoker (1957), Fenzl (1965) and Liggett and Woolhiser (1967) which clearly show that difference schemes may be used in practice with mesh increments outside the range to which existing numerical analysis methods apply. Simple examples follow in which instability may obviously arise from the non-homogeneous terms in the difference equations even in a "theoretically convergent" solution, and it is pointed out that existing numerical analysis does not comprehend the possibility of a physically unstable flow.

It is possible to resolve this discrepancy between theory and practice by linearising the difference equations by a new, more rational method before applying the Fourier series stability analysis, as is suggested by the introductory example in Chapter 4. Instead of assuming that the mesh increments are small, this new method produces linearised equations by assuming only that the variation in the solution is small, relative to the solution, across each mesh increment. Such equations are shown to be amenable to standard Fourier series stability analysis with the difference that physical stability criteria are now involved, together with factors reflecting the effects of the non-homogeneous terms. The results of this stability analysis show that the von Neumann necessary condition for stability is not also sufficient, at least for practical stability requirements, as stability errors may still grow rapidly even if this condition is satisfied. The bound of the amplification factor, by which stability errors are multiplied at each time step, is therefore set at the intuitively acceptable value of unity.

Finally in Chapter 4, it is shown that the "Courant condition" is still prominent in the newly derived stability criteria, and may normally be applied with adequate accuracy prior to a solution provided that the maximum discharge in that solution can be estimated approximately.

The extended Fourier series stability analysis is applied to a general case in Chapter 5 and the assumptions basic to the method are re-examined. Three are found to be virtually indispensable but not unduly restrictive - first, that the variations in the solution are "small" as discussed above; second, that the general difference scheme comprehends only one cycle, two level schemes on a regular mesh; and third, that the channel behaves essentially as a "quasi-simple" channel, which should include most regular or mildly irregular channels met in practice In particular, the commonly occurring trapezoidal channels are shown to be quasi-simple.

Physical stability criteria given by Vedernikov and Escoffier and Boyd are then shown to coincide for quasi-simple channels, where they are expressible in a particularly simple form.

The remainder of Chapter 5 is mainly devoted to the application of the extended stability analysis to the general difference scheme, the final three sections giving stability criteria for the general difference scheme in three special cases. The first of these is the case in which the mesh increments tend to zero, giving stability criteria from which known Richtmyer stability criteria for individual difference schemes are immediately recoverable. This is shown for one explicit and one implicit scheme. The second special case occurs when both the mesh increments and the variations between individual errors along any time line are small. In this case stability criteria are derived which depend only on the flow parameters. As these hold for infinitesimal mesh increments they must be criteria associated with the partial differential equations themselves, i.e. physical stability criteria. The third special case assumes no variation between individual errors along any time line, when two more stability criteria apply for all mesh sizes.

Before these last two special cases can be fully examined, the formulation of the non-homogeneous terms in the general difference equations must be clarified. Chapter 6 accordingly begins with a further assumption that these terms are all formulated at the same interior point of the difference scheme, which permits the identification of a number of factors in the stability analysis with dimensionless physical parameters. These are compared in magnitude for commonly occurring flows and the general physical stability criteria obtained in Chapter 5 are then translated into physical terms.

Complete stability analysis is more formidable than the study of selected special cases, as no assumptions can be made about the variation of errors along any time line, while the mesh increments cannot be taken as small. The generality of the stability analysis is therefore reduced in Chapter 6 by two more assumptions.

First, the space differences are assumed to be evaluated at the backward time line at which the solution is known, which eliminates fully implicit difference schemes. Second, the channel is taken to be near prismatic. The stability analysis then follows the pattern set out in the example of Chapter 4 until a general expression is obtained for the eigenvalues of the amplification matrix.

This expression incorporates complete stability criteria for a number of simple difference schemes, four of which are selected for further study. First the Lax "staggered" scheme is discussed briefly and it is shown that the same criteria are obtainable as by the introductory example of Chapter 4 which dealt with the same scheme. Then the "Unstable" scheme is discussed and empirical experience that this scheme may be stable if lateral inflows are present is confirmed by the form of the stability criteria. Next the rectangular scheme analysed by Courant, Isaacson and Rees and tried by Stoker is analysed and is shown to have poor stability properties for finite time increments, due entirely to the formulation of the non-homogeneous terms at the backward time line. The "Semi-Explicit" shceme, which is identical to that suggested by Courant, Isaacson and Rees except that the non-homogeneous terms are evaluated at the forward time line, is then shown to have stability properties for all time increments identical to those given for infinitesimal time increments by Courant, Isaacson and Rees. In other words, the sole numerical stability criterion governing the Semi-Explicit scheme is the "Courant condition".

The advantages and disadvantages of characteristic, implicit and explicit difference schemes are discussed in Chapter 7, with the conclusion that the Semi-Explicit scheme on a regular net has the best combination of simplicity, directness, speed and stability of any considered. Algorithms suitable for the computation of solutions by means of the Semi-Explicit scheme are then discussed, and it is shown that this scheme can be made reliable even when the velocity or depth in the solution becomes small, when our linearising assumption, that the variations in the solution are small compared with the solution itself, is invalid.

The predictions of the stability analysis of Chapters 5 amd 6 are then tested by an experimental comparison between the stability properties of the Courant, Isaacson and Rees scheme and those of the Semi-Explicit scheme. In all cases the stability analysis is found to give conservative but qualitatively accurate predictions. The substantial advantages of the Semi-Explicit scheme for large time steps is confirmed, but a slightly unexpected result is the increasing conservatism of the analytical stability criteria as the Froude Number was increased in the experiments, and the tests also showed that no sudden changes in the stability of the solution appeared as the Froude Number passed through unity. However these two effects should not be relied on as the stability experiments dealt only with uniform flow.

The final conclusion of Chapter 7 was that of all the difference schemes discussed the Semi-Explicit scheme is the best for general use.

An alternative approach to the direct numerical solution of the Overland Flow equations is the solution of a simplified form of overland flow equations. If it is assumed that acceleration and extra momentum terms are an unimportant part of the equation of motion the overland flow equations reduce to a single equation with depth (or stage) as a single dependent variable. The diffusion solution given by Hayami (1951) is outlined in Appendix F and it is shown in Chapter 8 that this solution may usefully be extended to apply to moderately irregular natural channels as well as the wide prismatic rectangular channels assumed by Hayami. The main requirement of this extended solution is some measure of the steady state stage preceding and following a disturbance which passes down the channel after being introduced at the upstream end. The extended solution is then shown to predict that at a given time from the initiation of such a disturbance, the change in stage at a given point on the channel will bear a constant ratio to the difference between the initial and final steady state stages at that point.

This prediction is tested experimentally in Chapter 9. One flood wave in the Clutha River is used as a calibration flood to permit stage/time curves to be observed along the river. Dimensionless graphs of the extended Hayami solution are used to produce curves of change in stage (as a proportion of final change in stage) versus time which provide the best fit of diffusion solutions to the stage/time curves measured at each observation point. A second, different flood wave is then introduced and the measured initial and final steady states are used with the fitted diffusion solution (relative change in stage)/time curves to predict the new stage/time curves at each observation point. A good fit is obtained between the predicted and observed stage/time records.

Diffusion methods therefore appear promising, particularly in describing a flow subject to artificial upstream control, as in this case the actual boundary condition approximates the convenient case of a series of steady states separated by sudden step changes.

## 10.2 Principal Conclusions

1. Overland Flow equations may be developed and solved without introducing the common assumptions that the flow is near horizontal, that the velocity is uniform over any cross section, that the lateral inflow is small, and that the channel is prismatic; and without introducing a "bed slope".
2. Existing solution theory, while forming a useful background,

fails to deal adequately with the non-homogeneous terms in overland flow equations. As such terms dominate the solutions of most overland flows, existing solution theory does not fulfil practical requirements.

3. This discrepancy between theory and practice in stability properties can be resolved by substituting for the usual assumption that the mesh increments are small the more axiomatic requirement that the difference solution does not vary substantially across a single mesh increment. Fourier series stability analysis using this requirement to linearise the difference equations then includes the non-homogeneous terms, and successfully predicts experimental stability properties which do not conform to the predictions of existing stability analysis.

4. Known criteria for physical flow stability appear in this new Fourier series stability analysis, and these can be expressed in the simple form of (5-9). These criteria are not greatly affected by moderate width variations and lateral inflows.

5. The numerical stability of the Semi-Explicit scheme (that proposed by Courant, Isaacson and Rees (1952), except that the non-homogeneous terms are evaluated at the forward time line) is shown by the new Fourier series analysis to be governed by the Courant condition alone, for all magnitudes of the mesh increments.

6. For general use, the Semi-Explicit scheme has demonstrable advantages over all other difference schemes considered, but careful programming is still required to obtain the best results

from the scheme.

7. An efficient, stable time step for the Semi-Explicit scheme can usually be chosen prior to a solution on the basis of an approximate estimate of the maximum discharge in that solution.
8. The diffusion analogy solution provides a simple, rapid reasonably accurate means of predicting the rate of propagation of a flood wave resulting from a sudden increase in discharge into the upstream end of the channel. This solution is sufficiently promising to warrant further investigation.
9. The results of this thesis show that a dependable general purpose numerical solution is available for many overland flow problems, but in special cases simplified solutions may give adequate results with considerably less effort.

## REFERENCES

Abbott, M.B. "An Introduction to the Method of Characteristics," (London: Thames and Hudson, 1966).

Abbott, M.B. and F. Ionescu. "On the Numerical Computation of Nearly Horizontal Flows," Journal of Hydraulic Research (IAHR) Vol.5 (1967) No.2 pp.97-117.

Aitken, A.C. "Determinants and Matrices," (Edinburgh: Oliver and Boyd, 1956).

Anon. "Fortran IV Library Subprograms," IBM Systems Reference Library, Form C28-6596-4 (1968).

Chow, V.T. "Open-Channel Hydraulics," (New York: McGraw-Hill Book Company, Inc., 1959).

Courant, R. and K.O. Friedrichs. "Supersonic Flow and Shock Waves," (New York: Interscience Publishers, Inc., 1967).

Courant, R., E. Isaacson and M. Rees. "On the Solution of Nonlinear Hyperbolic Differential Equations by Finite Differences," Comm. Pure and App. Maths, Vol.V (1952) pp.243-255.

Courant, R. and P. Lax. "On Nonlinear Partial Differential Equations with Two Independent Variables," Comm. Pure and App. Maths, Vol.II (1949) pp.255-273.

Escoffier, F.F. and M.B. Boyd. "Stability Aspects of Flow in Open Channels," Proc. ASCE, Vol.88, No.HY6 (November 1962), pp.145-166.

Fenzl, R.N. Discussion on "Computer Analysis of Overland Flow" by Morgali and Linsley (see below), Proc. ASCE, Vol.91, No. HY6 (November 1965), pp.224-227.

Hayami, S. "On the Propagation of Flood Waves," Bulletin No.1, Disaster Prevention Research Institute, Kyoto University, Japan (December 1951).

Henderson, F.M. "Open Channel Flow," (New York: The MacMillan Company, 1966).

Lamb, H. "Hydrodynamics," (New York: Dover Publications, 1945).

Lax, P.D. "Nonlinear Hyperbolic Equations," Comm. Pure and App. Maths, Vol.VI (1953) pp.231-258.

Lax, P.D. "Weak Solutions of Nonlinear Hyperbolic Equations and their Numerical Computation," Comm. Pure and App. Maths, Vol.VII (1954) pp.159-193.

Liggett, J.A. and D.A.Woolhiser. "Difference Solutions of the Shallow-Water Equations," Proc.ASCE, Vol.93, No. EM2 (April 1967), pp.39-71.

Liggett, J.A. and D.A.Woolhiser. Discussion on "A Solution to the Runoff Problem," Proc. ASCE, Vol.94, No.HY6 (November 1968), pp.1578-1582.

Lighthill, M.J. and G.B.Whitham. "On Kinematic Waves : I - Flood Movement in Long Rivers," Proc. Roy.Soc.(London), (A), Vol.229, No.1178 (May 1955), pp.281-316.

Milne-Thomson, L.M. "Theoretical Hydrodynamics" (London: MacMillan and Co.Ltd., 1960).

Morgali, J.R. and R.K.Linsley. "Computer Analysis of Overland Flow," Proc. ASCE, Vol.91, No.HY3 (May 1965), pp.81-100.

Richtmyer R.D. "Difference Methods for Initial-Value Problems," (New York: Interscience Publishers Inc., 1957).

Rouse, H.(ed.), "Engineering Hydraulics" (New York: John Wiley and Sons, Inc., 1950). Reference made to: Chapter I "Fundamental Principles of Flow" by H.Rouse. Chapter XI "Wave Motion" by G.H.Keulegan.

Sarma, K.V.N. and S.R.Sasikanth. "Evaluation of Manning's n for Steady Non-Uniform Flows" 2nd Australasian Conf. on Hydraulics and Fluid Mechanics, Auckland, November 1965, pp.A177-A186.

Smith, G.D. "Numerical Solution of Partial Differential Equations" London: Oxford University Press, 1965).

Stoker, J.J. "Water Waves" (New York: Interscience Publishers Inc., 1957).

Terzidis, G and T.Strelkoff. Discussion on "Computer Analysis of Overland Flow" by Morgali and Linsley (see above), Proc. ASCE, Vol.91, No.HY6 (November 1965), pp.236-238.

Thomas, I.E. and P.R.Wormleaton. "Flood Routing Using a Convective Diffusion Model," Civil Engineering and Public Works Review, Vol.65, No.764 (March 1970), pp.257-259.

Vreugdenhil, C.B. Discussion on "Difference Solutions of the Shallow-Water Equation" by Liggett and Woolhiser (see above), Proc. ASCE, Vol.94, No.EM1 (February 1968), pp.334-339.

## APPENDIX A. RICHTMYER STABILITY

Our principal reference in this Appendix is Richtmyer (1957), in which the possibilities of the Fourier series method of investigating stability are explored. This method is strictly applicable only to a restricted class of problems, in which (Richtmyer (1957), Chapter I) the differential equations are linear, all coefficients are constant, and the boundary conditions are equivalent to reflection and periodicity conditions. Thus the use of the method on nonlinear equations is not strictly valid. However, in many nonlinear equations, especially quasi-linear equations (Section 3.4), the nonlinear coefficients in the equations vary slowly in a continuous manner, so that in a small region of the solution they may be regarded as constants. Hence these equations may be regarded as "locally linear". Thus we should gain valuable insight into the stability properties of many nonlinear solutions by applying Fourier series methods to a small region, although the stability properties will usually be expressed as a function of the "locally constant" coefficients.

Although Richtmyer does not deal with hydrodynamic systems, he examines compressible fluid flows in Chapter X. Analysis of this problem leads to an hyperbolic system of quasi-linear partial differential equations analagous to the homogeneous form of the Overland Flow equations. We shall therefore follow the salient points of Richtmyer's discussion, presenting as an example an investigation into the application of the rectangular finite difference scheme, recommended by Stoker (1957) p.477, to the homogeneous form of the Overland Flow equations. This rectangular scheme is the one investigated by Courant, Isaacson and Rees (1952), so it will be called the CIR rectangular scheme in this thesis.

For convenience we assume that the channel is prismatic simple, so that from (3-28) $w = 2kc$. We shall also term the characteristic along which $ds/dt - V = +c$ "forward", and that along which $ds/dt - V = -c$ "backward" (see equation (3-17)).

Hence the homogeneous equation along the forward characteristic becomes, from (3-25)

$$V_t + 2kc_t + (V + c)(V_s + 2kc_s) = 0 \tag{A-1}$$

while along the backward characteristic, from (3-26)

$$V_t - 2kc_t + (V - c)(V_s - 2kc_s) = 0 \tag{A-2}$$

We assume that we are discussing a small region of the solution in which the coefficients V and c can be regarded as "locally constant".

Now we define a rectangular lattice on the s-t plane such that the grid lines are described by $s = (0, 1, 2, \ldots\ a-1, a, a+1, \ldots\ K_a)\Delta s$ and $t = (0, 1, \ldots\ b, b+1, \ldots\ K_b)\Delta t$. We use $V_a^b$, $c_a^b$ to denote the values $V(a\Delta s, b\Delta t)$, $c(a\Delta s, b\Delta t)$ respectively. The CIR rectangular scheme uses the finite difference expressions

$$V_t = \frac{V_a^{b+1} - V_a^b}{\Delta t} \qquad c_t = \frac{c_a^{b+1} - c_a^b}{\Delta t} \tag{A-3}$$

Along the forward characteristic, that is in equation (A-1), we use

$$V_s = \frac{V_a^b - V_{a-1}^b}{\Delta s} \qquad c_s = \frac{c_a^b - c_{a-1}^b}{\Delta s} \tag{A-4}$$

and in equation (A-2), along the backward characteristic

$$V_s = \frac{V_{a+1}^b - V_a^b}{\Delta s} \qquad c_s = \frac{c_{a+1}^b - c_a^b}{\Delta s} \tag{A-5}$$

as these differences correspond to the derivatives applying along the forward and backward characteristics passing though the line $s = a\Delta s$, provided $V - c$ is negative. If we test for consistency by the method outlined in Section 3.6, we find that the difference equations differ from the differential equations by terms $O(\Delta s, \Delta t)$ as $\Delta s$ and $\Delta t$ tend to zero. Thus consistency is satisfied by the CIR rectangular finite difference scheme. We must now analyse the stability properties of the scheme.

Substituting (A-3) and (A-4) in (A-1), and (A-3) and (A-5) in (A-6), then eliminating first $V_a^{b+1}$, then $c_a^{b+1}$, we get

$$c_a^{b+1} = c_a^b - D_v c\frac{\Delta t}{\Delta s} - D_c V\frac{\Delta t}{\Delta s} \tag{A-6}$$

$$V_a^{b+1} = V_a^b - 2kD_v V\frac{\Delta t}{\Delta s} - 2kD_c c\frac{\Delta t}{\Delta s} \tag{A-7}$$

where $$D_v = \frac{V_{a+1}^b - V_{a-1}^b}{4k} + c_a^b - \tfrac{1}{2}(c_{a+1}^b + c_{a-1}^b) \tag{A-8}$$

and $$D_c = \frac{c_{a+1}^b - c_{a-1}^b}{2} + \frac{V_a^b - \tfrac{1}{2}(V_{a+1}^b + V_{a-1}^b)}{2k} \tag{A-9}$$

We let $\Delta t$ and $\Delta s$ tend to zero together, maintaining a constant ratio $\Delta t/\Delta s = r$. Thus we can replace (A-6) and (A-7) with their equations of first variation by substituting $V + \mathring{V}_a^b$,

$c + \mathring{c}_a^b$ for $V_a^b$, $c_a^b$ etc, where $\mathring{V}$, $\mathring{c}$ are small quantities of the first order, while V and c are zero order "local constants". Note that $D_v$ and $D_c$ are terms of the first order as all the zero order quantities cancel in (A-8) and (A-9). Discarding quantities of the second and higher orders, the equations of first variation are

$$\mathring{c}_a^{b+1} = \mathring{c}_a^b - D_v cr - D_c Vr \tag{A-10}$$

$$\mathring{V}_a^{b+1} = \mathring{V}_a^b - 2kD_v Vr - 2kD_c cr \tag{A-11}$$

The coefficients c and V in (A-6) and (A-7) are evaluated as some function of the solutions at grid points in an actual finite difference solution, but any variation quantity associated with those grid points contributes only second order quantities to the equations (A-10) and (A-11) because the coefficients are multiplied by the first order quantities $D_v$ and $D_c$. Hence the actual method of evaluating the coefficients c and V is of secondary importance in this stability analysis.

Now we let the vector u represent $\begin{bmatrix} \mathring{c} \\ \mathring{V} \end{bmatrix}$ and express u as a Fourier series

$$u\ (s,t) = \sum_{p=-\infty}^{\infty} \overset{*}{u}\ (p,t)e^{ips} \tag{A-12}$$

where $\overset{*}{u}$ is the finite complex Fourier transform of u. Remembering $\overset{*}{u}(p,t) = \overset{*}{u}(p,b\Delta t)$, we denote $\overset{*}{u}(p,t)$ by $\overset{*}{u}^b(p)$.

Now as $\overset{*}{u}^b(p)$ is independent of s, we need only investigate the amplification, for all p, of $\overset{*}{u}$ as t increases along a typical line $s = a\Delta s$, in order to evaluate the stability properties of a finite difference scheme. Remembering that $\Delta s = \Delta t/r$, we may substitute from (A-12) into (A-10) and (A-11) and rearrange to find an equation system which may be written

$$\overset{*}{u}^{b+1} = M_a(\Delta t, p)\overset{*}{u}^{b} \tag{A-13}$$

$M_a$ is called the amplification matrix by Richtmyer, who derives a generalized amplification matrix for simple linear initial value problems in Chapter IV of his book (1957). He shows that in such problems a necessary and sufficient stability condition requires that n operations of $M_a$ produce a bounded result, where $0 \leqslant n \leqslant T/\Delta t$ and T is the time over which the solution is carried. Thus, because each eigenvalue $\lambda_j$ of $M_a$ is less than or equal to the bound of $M_a$, a necessary condition for stability is that each $(\lambda_j)^n$ must be bounded. This means that we require a constant $C_1$ to exist such that $|\lambda_j| \leqslant C_1^{\Delta t/T}$.

We can assume that $C_1 \geqslant 1$, and hence we have the condition

$$|\lambda_j| \leqslant 1 + O(\Delta t) \tag{A-14}$$

which is the von Neumann necessary condition for stability.

Richtmyer concludes Chapter IV by stating that in many (linear) cases in which he has been able to investigate full conditions for stability, the von Neumann condition has turned out to be sufficient as well as necessary. Hence if we satisfy the von Neumann condition in all local areas of our non linear solution, it is likely that the finite difference scheme is stable, at least as $\Delta t$ tends to zero.

We shall therefore investigate the behaviour of $\lambda_j$ as

$\Delta t$ tends towards zero.

Expressing u(s,t) as a Fourier series (A-12) we note that the factor $e^{ipa\Delta s}$ may be cancelled throughout, so that (A-10) becomes, after substituting for $D_v$ and $D_c$

$$\overset{*}{c}^{b+1} = \left[1 - c\frac{\Delta t}{\Delta s}(1 - \tfrac{1}{2}(e^{ip\Delta s} + e^{-ip\Delta s})) - \frac{V}{2}\frac{\Delta t}{\Delta s}(e^{ip\Delta s} - e^{-ip\Delta s})\right]\overset{*}{c}^{b}$$
$$- \left[\frac{c\Delta t}{4k\Delta s}(e^{ip\Delta s} - e^{-ip\Delta s}) + \frac{V\Delta t}{2k\Delta s}(1 - \tfrac{1}{2}(e^{ip\Delta s} + e^{-ip\Delta s}))\right]\overset{*}{V}^{b} \qquad \text{(A-15)}$$

using the notation $\overset{*}{c}^{b}$ for the Fourier transform of $\mathring{c}_a^b$, and so on.

Now p may vary to any integral value, but we achieve the same variation by setting $p\Delta s = \theta$ and allowing $\theta$ to vary continuously between $-\pi$ and $\pi$. In this way we can explore the principal values of $M_a$. Thus

$$1 - \tfrac{1}{2}(e^{ip\Delta s} + e^{-ip\Delta s}) = 1 - \cos\theta \qquad \text{(A-16)}$$

$$e^{ip\Delta s} - e^{-ip\Delta s} = 2i\sin\theta \qquad \text{(A-17)}$$

We also introduce the substitutions

$$W = \frac{V\Delta t}{\Delta s} \qquad Z = \frac{c\Delta t}{\Delta s} \qquad \text{(A-18)}$$

Therefore (A-15) becomes

$$\overset{*}{c}^{b+1} = (1 - Z(1 - \cos\theta) - iW\sin\theta)\overset{*}{c}^{b} - (W(1 - \cos\theta) + iZ\sin\theta)\frac{\overset{*}{V}^{b}}{2k} \qquad \text{(A-19)}$$

Likewise (A-11) gives

$$\overset{*}{V}^{b+1} = -(W(1 - \cos\theta) + iZ\sin\theta)2k\overset{*}{c}^{b} + (1 - Z(1 - \cos\theta) - iW\sin\theta)\overset{*}{V}^{b} \qquad \text{(A-20)}$$

Referring to (A-13) we see that for this example

$$M_a = \begin{bmatrix} 1 - Z(1-\cos\theta) - iW\sin\theta & -\frac{1}{2k}(W(1-\cos\theta) + iZ\sin\theta) \\ -2k(W(1-\cos\theta) + iZ\sin\theta) & 1 - Z(1-\cos\theta) - iW\sin\theta \end{bmatrix} \quad \text{(A-21)}$$

We find the eigenvalues from the equation $|M_a - \lambda I| = 0$,

viz $$\lambda_{1,2} = 1 - r\left[(c \pm V)(1 - \cos\theta \pm i\sin\theta)\right] \quad \text{(A-22)}$$

Thus $$|\lambda_1| = \left[(1 - r(c+V)(1-\cos\theta))^2 + (r(c+V)\sin\theta)^2\right]^{\frac{1}{2}}$$

$$= \left[1 - 2r(c+V)(1-\cos\theta)(1 - r(c+V))\right]^{\frac{1}{2}} \quad \text{(A-23)}$$

Now $(1-\cos\theta)$ is positive, so that if c+V is negative, $|\lambda_1|$ exceeds unity for all $\theta(\neq 0)$. If c+V is positive $|\lambda_1|$ exceeds unity for all $\theta$ if 1 - r(c+V) is negative.

Similarly

$$|\lambda_2| = \left[1 - 2r(c-V)(1-\cos\theta)(1 - r(c-V))\right]^{\frac{1}{2}} \quad \text{(A-24)}$$

and by similar reasoning $|\lambda_2|$ exceeds unity only if c-V is negative or if 1 - r(c-V) is negative.

Thus the von Neumann necessary condition for stability (A-14) requires

$$c - |V| \geq 0 \quad \text{(A-25)}$$

and $$(c + |V|)\frac{\Delta t}{\Delta s} \leq 1 \quad \text{(A-26)}$$

If we compare these criteria with the Courant condition (Section 3.5), we see that the domain of dependence of the point $P(a\Delta s,(b+1)\Delta t)$ used by this difference scheme is the range $(a-1)\Delta s \leq s \leq (a+1)\Delta s$ at $t = b\Delta t$. If we allow $\Delta t/\Delta s$ to increase until (A-26) is not satisfied, the point P then has an actual domain of dependence as determined by the differential equations which is outside the range used by the difference scheme. Hence in this case a disturbance at time $t = b\Delta t$ from outside the range $(a-1)\Delta s \leqslant s \leqslant (a+1)\Delta s$ may influence P, but the difference equations do not comprehend this possibility. Thus criterion (A-26) is a quantitative expression of the Courant condition, which corroborates the Fourier series method of stability analysis. Condition (A-25) reflects Courant's comment that the choice of difference quotients should preserve the domain of dependence, because in using (A-4) and (A-5) we made the tacit assumption that the forward characteristic through P also passed between $((a-1)\Delta s,\ b\Delta t)$ and $(a\Delta s,\ b\Delta t)$, while similarly we took it that the backward characteristic through P also passed between $(a\Delta s,\ b\Delta t)$ and $((a+1)\Delta s,\ b\Delta t)$. Even if (A-26) is satisfied, this is true only if the flow is subcritical, and (A-25) simply expresses this restriction.

Clearly Fourier series stability analysis is a useful tool for investigating the behaviour of finite difference solutions of the Overland Flow equations, at least as $\Delta t$ and $\Delta s$ tend to zero.

## APPENDIX B. FULL EVALUATION OF THE SQUARE ROOT

This appendix proves equation (6-77) given (6-71), (6-72) and (6-76). Variables not defined here are as defined previous to Section 6.7 in the thesis.

We set

$$\Omega = (\epsilon X + \zeta \Phi_2)^2 \tag{B-1}$$

Subtracting (B-1) from (6-63) we have

$$(L_1{}^2\Delta t^2 - \epsilon^2)X^2 + 2(L_2\Delta t\Gamma_1 - \epsilon\zeta)X\Phi_2 + (\Gamma_2^2 - \zeta^2)\Phi_2{}^2 = 0 \tag{B-2}$$

Now X and $\Phi_2$ are complex variables, so we can say

$$X^2 = R_1 + i \sin\theta \, I_1$$

$$X\Phi_2 = R_2 + i \sin\theta \, I_2$$

$$\Phi_2{}^2 = R_3 + i \sin\theta \, I_3$$

where $R_{1,2,3}$ and $I_{1,2,3}$ are all real variables. We have written the imaginary parts as such because they can arise only from the imaginary part of $e^{i\theta}$ which disappears if $\sin\theta = 0$.

Thus (B-2) can be written as two equations

$$(L_1{}^2\Delta t^2 - \epsilon^2)R_1 + 2(L_2\Delta t\Gamma_1 - \epsilon\zeta)R_2 + (\Gamma_2^2 - \zeta^2)R_3 = 0 \tag{B-3}$$

$$(L_1{}^2\Delta t^2 - \epsilon^2)I_1 + 2(L_2\Delta t\Gamma_1 - \epsilon\zeta)I_2 + (\Gamma_2^2 - \zeta^2)I_3 = 0 \tag{B-4}$$

where we have divided the imaginary part by $\sin\theta$ throughout, assuming $\sin\theta \neq 0$. Because we need two equations to solve our two unknowns $\epsilon$ and $\zeta$, we shall require (B-4) to hold for $\sin\theta = 0$, as an arbitrary relationship between $\epsilon$ and $\zeta$ is needed in this case. This is reasonable as (B-4) must hold as $\sin\theta$ approaches zero.

We are assuming that $L_1$, $L_2$, $\Gamma_1$ and $\Gamma_2$ are real, which follows if $G_1$, $G_2$, $G_3$, $\varkappa_1$ and $\alpha_1$ are real. The only possibilities of $\varkappa_1$ and $\alpha_1$ being complex are if the averages they represent are not centred around the line $a\Delta s$, or if a difference term is included in the averages, and these are discounted. Thus from (B-2) we suppose that $\epsilon^2$, $\epsilon\zeta$ and $\zeta^2$ are also real, and we prove in a moment that such real parameters exist.

Subtracting (B-4) x $R_1$ from (B-3) x $I_1$ gives

$$2(L_2\Delta t\Gamma_1 - \epsilon\zeta)(I_1R_2 - I_2R_1) + (\Gamma_2^2 - \zeta^2)(I_1R_3 - I_3R_1) = 0 \tag{B-5}$$

Put
$$N = \frac{I_1R_3 - I_3R_1}{I_1R_2 - I_2R_1} \tag{B-6}$$

$$X = R_4 + i \sin\theta\ I_4 \tag{B-7}$$

$$\Phi_2 = R_5 + i \sin\theta\ I_5 \tag{B-8}$$

where $R_{4,5}$ and $I_{4,5}$ are real.

Hence
$$R_1 = R_4^2 - I_4^2 \sin^2\theta$$

$$I_1 = 2I_4R_4$$

$$R_2 = R_4R_5 - I_4I_5 \sin^2\theta$$

$$I_2 = R_4I_5 + I_4R_5$$

$$R_3 = R_5^2 - I_5^2 \sin^2\theta$$

$$I_3 = 2I_5R_5$$

So that
$$I_1R_3 - I_3R_1 = 2I_4R_4(R_5^2 - I_5^2\sin^2\theta) - 2I_5R_5(R_4^2 - I_4^2\sin^2\theta)$$
$$= 2(R_4R_5 + I_4I_5 \sin^2\theta)(I_4R_5 - I_5R_4)$$

Similarly $I_1R_2 - I_2R_1 = (R_4^2 + I_4^2\sin^2\theta)(I_4R_5 - I_5R_4)$

$$I_2R_3 - I_3R_2 = (R_5^2 + I_5^2\sin^2\theta)(I_4R_5 - I_5R_4)$$

Therefore, from (B-6)

$$N = \frac{2(R_4R_5 + I_4I_5\sin^2\theta)}{|X|^2} \tag{B-9}$$

$|X|$ will be zero only if $R_4$ and $I_4 \sin\theta$ are simultaneously zero and this is extremely unlikely. We therefore assume $|X| \neq 0$.

Thus we can say, from (B-5)

$$\epsilon = \frac{2L_2\Delta t\Gamma_1 + N(\Gamma_2^2 - \zeta^2)}{2\zeta} \tag{B-10}$$

$$\epsilon^2 = \frac{1}{4\zeta^2}\left[N^2\zeta^4 - 2N\zeta^2(N\Gamma_2^2+2L_2\Delta t\Gamma_1) + (N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)^2\right]$$

Substituting in (B-3) and multiplying by $\zeta^2$

$$L_1^2\Delta t^2\zeta^2R_1 - \tfrac{1}{4}R_1N^2\zeta^4 + \tfrac{1}{2}R_1N\zeta^2(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)$$
$$- \tfrac{1}{4}R_1(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)^2 + (\zeta^2-\Gamma_2^2)(NR_2-R_3)\zeta^2 = 0$$

Rearranging

$$(NR_2-R_3 - \tfrac{1}{4}N^2R_1)\zeta^4 + (L_1^2\Delta t^2R_1 + \tfrac{1}{2}NR_1(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)$$
$$- \Gamma_2^2(NR_2-R_3))\zeta^2 - \tfrac{1}{4}R_1(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)^2 = 0 \tag{B-11}$$

We have apparently taken $I_1$ and $R_1$ to be non-zero in this manipulation. By our assumption that $|X| \neq 0$, $I_1$ and $R_1$ cannot simultaneously be zero. If $I_1$ is zero then (B-10) follows directly from (B-3), while if $R_1$ is zero then (B-10) follows from (B-3), but in this case we substitute (B-10) into (B-4) to get an equation identical to (B-11) except $I_1$, $I_2$, $I_3$

replace $R_1$, $R_2$, $R_3$ respectively.

Now $$NR_2 - R_3 = \frac{(I_2R_3 - I_3R_2)}{I_1R_2 - I_2R_1} R_1 = \frac{|\Phi_2|^2}{|X|^2} R_1$$

and $$NI_2 - I_3 = \frac{|\Phi_2|^2}{|X|^2} I_1$$

So that if $R_1 \neq 0$ we divide (B-11) by $R_1$ throughout to give

$$\left[\frac{|\Phi_2|^2}{|X|^2} - \tfrac{1}{4}N^2\right]\zeta^4 + \left[L_1{}^2\Delta t^2 + \tfrac{1}{2}N(N\Gamma_2{}^2 + 2L_2\Delta t\Gamma_1) - \Gamma_2{}^2 \frac{|\Phi_2|^2}{|X|^2}\right]\zeta^2 - \tfrac{1}{4}(N\Gamma_2{}^2 + 2L_2\Delta t\Gamma_1)^2 = 0 \qquad \text{(B-12)}$$

If $R_1 = 0$ we replace $R_1$, $R_2$, $R_3$ by $I_1$, $I_2$, $I_3$ in (B-11) and division by $I_1$ again gives (B-12).

Now $$\frac{|\Phi_2|^2}{|X|^2} - \tfrac{1}{4}N^2 = \left[\frac{(I_4R_5 - I_5R_4)\sin\theta}{|X|^2}\right]^2 = J_1{}^2\sin^2\theta \qquad \text{(B-13)}$$

where $J_1$ is real.

Thus (B-12) is a quadratic in $\zeta^2$, provided $\sin\theta \neq 0$, with real roots $\zeta_1{}^2$, $\zeta_2{}^2$ where

$$\zeta_1{}^2\zeta_2{}^2 = -\frac{(N\Gamma_2{}^2 + 2L_2\Delta t\Gamma_1)^2}{4J_1{}^2\sin^2\theta}$$

Hence one of $\zeta_1{}^2$, $\zeta_2{}^2$ is positive and the other negative, proving that we can select $\zeta$ to be either real or imaginary.

$\zeta_1$ and $\zeta_2$ are related by

$$\zeta_1 = \pm\frac{i(N\Gamma_2{}^2 + 2L_2\Delta t\Gamma_1)}{2J_1\sin\theta\,\zeta_2} \qquad \text{(B-14)}$$

From (B-10) $$\epsilon_{1,2} = \frac{2L_2\Delta t\Gamma_1 + N\Gamma_2{}^2}{2\zeta_{1,2}} - \tfrac{1}{2}N\zeta_{1,2}$$

$$\epsilon_1 X + \zeta_1 \Phi_2 = \left[\frac{2L_2\Delta t\Gamma_1 + N\Gamma_2^2}{2\zeta_1} - \tfrac{1}{2}N\zeta_1\right]X + \zeta_1\Phi_2$$

$$= \mp\, iJ_1 \sin\theta \;\; X\zeta_2 + (\Phi_2 - \tfrac{1}{2}NX)\zeta_1$$

$$= \mp\, iJ_1 \sin\theta \;\; X\zeta_2 - iJ_1 \sin\theta \;\; X\zeta_1$$

$$= \pm\, \epsilon_2 X \pm (\tfrac{1}{2}N - iJ_1 \sin\theta)X\zeta_2$$

$$= \pm\,(\epsilon_2 X + \zeta_2\Phi_2) \tag{B-15}$$

which checks our algebraic manipulation so far as we used (B-10) (B-14), (B-7), (B-8), (B-9) and (B-13) in turn to derive (B-15). Note that from (B-10) $\epsilon_1$ is real if $\zeta_1$ is real and imaginary if $\zeta_1$ is imaginary, and similarly with $\epsilon_2$ and $\zeta_2$. This is consistent with our assumption that $\epsilon^2$, $\epsilon\zeta$ and $\zeta^2$ are all real.

Now (B-15) holds for $\sin\theta \neq 0$, but if $\sin\theta = 0$, (B-12) becomes, using (B-13)

$$J_2\zeta^2 = \tfrac{1}{4}(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)^2 \tag{B-16}$$

where $J_2 = L_1^2\Delta t^2 + \tfrac{1}{2}N(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1) - \Gamma_2^2 \dfrac{|\Phi_2|^2}{|X|^2}$

$$= \frac{1}{\Gamma_2^2}\Big[(\Gamma_2^2 L_1^2 - \Gamma_1^2 L_2^2)\Delta t^2 + \tfrac{1}{4}(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)^2 - \Gamma_2^4 J_1^2 \sin^2\theta\Big]$$

$$= \frac{1}{\Gamma_2^2}\Big[2k\ \det(M_1)(L_1^2 - L_2^2)\Delta t + \tfrac{1}{4}(N\Gamma_2^2 + 2L_2\Delta t\Gamma_1)^2 - \Gamma_2^4 J_1^2 \sin^2\theta\Big] \tag{B-17}$$

using (B-13), (6-75) and (6-62).

Thus when $\sin\theta = 0$, $J_2$ is positive from (6-76), (6-71) and (6-72), and therefore from (B-16) $\zeta$ is real for $\sin\theta = 0$. Because (B-10) holds for all $\theta$, $\epsilon$ is also real for $\sin\theta = 0$. Thus real values $\epsilon_r$ and $\zeta_r$ exist for all $\theta$, so that rearranging (B-12) using (B-13) and (B-17) gives

$$\left[\left(\frac{N\Gamma_2^2 + 2L_2\Delta t\Gamma_1}{2\Gamma_2}\right)^2 + \zeta_r^2 J_1^2 \sin^2\theta\right](\Gamma_2^2 - \zeta_r^2) = \frac{2k\ \det(M_1)}{\Gamma_2^2}(L_1^2 - L_2^2)\Delta t^2 \zeta_r^2 \qquad \text{(B-18)}$$

Also from (B-10) we have $2(L_2\Delta t\Gamma_1 - \epsilon_r\zeta_r) = N(\zeta_r^2 - \Gamma_2^2)$

so that dividing (B-3) by $R_1$ or (B-4) by $I_1$ gives

$$(L_1^2\Delta t^2 - \epsilon_r^2)\ |X|^2 = (\Gamma_2^2 - \zeta_r^2)|\Phi_2|^2 \qquad \text{(B-19)}$$

Hence (6-76) and (B-18) give

$$|\zeta_r| \leqslant |\Gamma_2| \qquad \text{(B-20)}$$

(B-19) and (B-20) give

$$|\epsilon_r| \leqslant |L_1\Delta t| \qquad \text{(B-21)}$$

and we can apply (B-20) and (B-21) to (B-1) to prove (6-77) as required.

## APPENDIX C. AN ALGORITHM FOR QUADRATIC SOLUTIONS

In this appendix an algorithm is developed for evaluating quadratic solutions which are ill conditioned in the way described in Section 7.4. Variables not defined here are defined in Sections 7.3 and 7.4.

We substitute

$$x = \frac{2p}{K_b^{\,2}} \tag{C-1}$$

And for $2x < 1$ we can write

$$-K_b + K_b(1 + 2x)^{\frac{1}{2}} = -K_b + K_b\left[1 + \tfrac{1}{2}(2x) + \frac{(\frac{1}{2})(-\frac{1}{2})}{2!}(2x)^2 + \ldots\ldots + \frac{(\frac{1}{2})(-\frac{1}{2})\ldots.(\frac{3-2n}{2})}{n!}(2x)^n + \ldots.\right]$$

$$= K_b\left[x + \frac{1(-1)}{2!}\,x^2 + \frac{1(-1)(-3)}{3!}\,x^3 + \ldots.. + (-n)^{n-1}\,\frac{1.3.5\ldots(2n-3)}{n!}\,x^n + \ldots.\right] \tag{C-2}$$

The series is convergent for $2x < 1$ and rapidly convergent for $2x$ small, when it is more conveniently expressed as a power series in $x$ as in (C-2). Now the use of the standard square root function will clearly result in a loss of one decimal place in working precision when $x = L$, where $L$ is $O(1/10)$, and this is a convenient point at which to switch to a truncated series based on (C-2). We write

$$x + \frac{1(-1)x^2}{2!} + \frac{(-1)(-3)}{3!}x^3 + \ldots + (-1)^{n-1}\frac{1.3.5.\ldots(2n-3)}{n!}x^n$$

$$= x - \frac{x^2}{2} + \frac{x^3}{2} - \frac{5}{8}x^4\Big[1 - \frac{7}{5}x + \frac{7}{5}.\frac{9}{6}x^2 - \frac{7}{5}.\frac{9}{6}.\frac{11}{7}x^3 + \ldots$$

$$+ (-1)^m \frac{7}{5}.\frac{9}{6}\ldots\frac{(2m+5)}{m+4}x^m \ldots\Big]$$

$$= x - \frac{x^2}{2} + \frac{x^3}{2} - hx^4 \qquad (C-3)$$

where h can be treated as constant if x is small. The relative error in so doing is

$$\epsilon = \frac{x^3\Big[\frac{5}{8}(1 - \frac{7}{5}x + \frac{7}{5}.\frac{9}{6}x^2 - \ldots) - h\Big]}{1 - \frac{x}{2} + \frac{x^2}{2} - \ldots} \qquad (C-4)$$

$$= x^3\Big[\frac{5}{8}(1 - \frac{7}{5}x + \frac{7}{5}.\frac{9}{6}x^2 - \ldots) - h\Big]\Big[1 - \frac{x}{2} + \frac{x^2}{2} - \ldots\Big]^{-1}$$

$$= x^3\Big[\frac{5}{8}(1 - \frac{7}{5}x + \frac{7}{5}.\frac{9}{6}x^2 - \ldots) - h\Big]\Big[1 + (\frac{x}{2} - \frac{x^2}{2} + \ldots)$$

$$+ (\frac{x}{2} - \frac{x^2}{2} + \ldots)^2 + \ldots\Big]$$

$$= x^3\Big[\frac{5}{8}(1 - \frac{7}{5}x + \frac{7}{5}.\frac{9}{6}x^2 - \ldots) - h\Big]\Big[1 + \frac{x}{2} - \frac{x^2}{4} + \ldots\Big] \qquad (C-5)$$

Thus for $x \leqslant L$ we can truncate the series arising from the denominator in (C-4) at 1, which gives a slight underestimate of $\epsilon$, or at $1 + \frac{x}{2}$, which gives a slight overestimate of $\epsilon$. The range is no more than about 5% if $x \leqslant L$, so that either procedure gives adequate accuracy for an error estimate. We therefore take

$$\epsilon \approx x^3\Big[\frac{5}{8}(1 - \frac{7}{5}x + \frac{7}{5}.\frac{9}{6}x^2 - \ldots) - h\Big] \qquad (C-6)$$

We wish to select the value of h which gives the minimum average relative error in the range $0 \leqslant x \leqslant L$. We therefore solve for h from

$$\frac{d}{dh}\left[\int_0^L \epsilon^2 dx\right] = 0$$

That is

$$\frac{d}{dh}\left[\frac{L^7}{7}h^2 - 2h.\frac{5}{8}\left(\frac{L^7}{7} - \frac{7}{5}\frac{L^8}{8} + \frac{7}{5}.\frac{9}{6}\frac{L^9}{9} - \ldots\right) + \frac{25}{64}\frac{L^7}{7} + \ldots\right] = 0$$

$$h = \frac{35}{8}\left(\frac{1}{7} - \frac{7}{5}\frac{L}{8} + \frac{7}{5}.\frac{9}{6}\frac{L^2}{9} - \ldots\right)$$

$$= 0.558 \quad \text{if } L = 0.1 \tag{C-7}$$

$$\frac{d^2}{dh^2}\left[\int_0^L \epsilon^2 dx\right] = \frac{2L^7}{7}$$ which is positive, so that (C-7) represents a minimum.

$$\frac{d\epsilon}{dx} = 3x^2\left[\frac{5}{8} - h - \frac{7}{6}x + \frac{35}{16}x^2 - \frac{33}{8}x^3 + \ldots\right] \tag{C-8}$$

$$\frac{d^2\epsilon}{dx^2} = 6x\left[\frac{5}{8} - h - \frac{7x}{6} + \frac{35}{16}x^2 - \frac{33}{8}x^3 + \ldots - \frac{7}{12}x + \frac{35}{16}x^2 - \frac{99}{16}x^3 + \ldots\right] \tag{C-9}$$

Ignoring the obvious minimum error when $x = 0$ we can say from (C-8) that a turning point exists when

$$\frac{5}{8} - h = \frac{7}{6}x - \frac{35}{16}x^2 + \frac{33}{8}x^3 + \ldots \tag{C-10}$$

and from (C-9) this is a maximum, assuming $x \leqslant L$. We call the solution of (C-10) $x_1$, so that

$$\epsilon_{max} = x_1^4\left[\frac{7}{24} - \frac{7}{8}x_1 + \frac{33}{16}x_1^2 - \ldots\right] \tag{C-11}$$

This is the maximum positive value of $\epsilon$, but a greater negative magnitude is possible in the range $x \leqslant L$ if $x_1 < L$, which will be the case if h is to minimize the average error in some way. As $d\epsilon/dx$ is negative for $x_1 < x \leqslant L$ in this case a minimum value of $\epsilon$ will occur at $x = L$ and if negative this is the minimum $\epsilon$ for the range $x \leqslant L$. Thus

$$|\epsilon| \leqslant \Big[x_1^{\,4}(\tfrac{7}{24} - \tfrac{7}{8}x_1 + \tfrac{33}{16}x_1^{\,2} - \ldots),$$
$$L^3(h - \tfrac{5}{8} + \tfrac{7}{8}L - \tfrac{21}{16}L^2 + \tfrac{33}{16}L^3 - \ldots .)\Big]_{max} \qquad \text{(C-12)}$$

For $L = 0.1$, $h = 0.558$ from (C-7) and thus, solving for $x_1$ in (C-10) by successive approximations, $x_1 = 0.064$.

Hence $|\epsilon| \leqslant \left[4.1 \times 10^{-6},\ 9.2 \times 10^{-6}\right]_{max}$ (C-13)

Thus if we use $h = 0.558$, $|\epsilon| \leqslant 10^{-5}$ even if we allow for the 5% underestimate of the error at $x = L$ which we introduced by truncating the denominator in (C-4). This is comparable with the error involved in the use of the standard square root function at $x = L$, because this function normally has a relative error of about $10^{-6}$ to $10^{-7}$, so that its use when $x = L$ introduces a relative error of $10^{-5}$ to $10^{-6}$.

For instance, the IBM System/360 computers used by the writer have a SQRT subprogram with a maximum relative error of $8.70 \times 10^{-7}$ (Anon (1968) p.45) which therefore gives an error, when $x = L$, of the same order as the use of (C-3) with $h = 0.558$.

Greater precision may be obtained from the series in (C-2) if required either by reducing L and suitably altering h or by retaining further terms in the series, but it is clear that this series expansion allows the rapid evaluation of the quadratic solution to any desired precision when x is small, increasing as x tends to zero, and hence we are able to eliminate ill conditioning from the solution of (7-26) without appreciably increasing computing time.

# APPENDIX D. PROGRAM FLOTS1



```
      INTEGER GETCLK
      DIMENSION DX(27),S(27),EM(27),F(27),VCDX(27),GSBV(27),CRP(27,2)
      COMMON C(9),V(9),CN(9),VN(9)
      CALL SETCLK
      DO 8 N1=1,3
      K=0
      Y=0.0016*25**N1
      DO 7 N2=1,3
      DXK=10.*3.162278**N2*Y**1.222222
      DO 7 N3=1,3
      SK=1.E-5*10**N3
      DO 7 N4=1,3
      K=K+1
      CALL IZETS1(DX,K,DXK,S,SK,EM,N4,CI,Y,VI,Q,F,VCDX,GSBV)
      DO 7 L=1,2
      CR=1.
      M=1
    1 CALL HEDTS1(L,CR,VCDX,K,GSBV,M,Y,DXK,SK,EM,F,VCTX,GSTV,FEMDT,GSDT,
     1DTX,CI,VI,Q)
      M=M+1
      ML=0
      J=1
    2 IF(L.EQ.2)GO TO 3
      CALL STRTS1(DTX,ML,FEMDT,GSDT)
      GO TO 4
    3 CALL BNTTS1(DTX,ML,GSDT,FEMDT)
    4 CALL TESTS1(J,ML,CI,VI,Y,EM,K,SK,&2)
      IF(CR.LT.1.)GO TO 5
      IF(CR.GT.1.)GO TO 6
      IF(ML.LT.10)CR=1.1
      IF(ML.GE.10)CR=0.95
      GO TO 1
    5 IF(CR.LE.0.05)GO TO 7
      IF(ML.GE.10)CR=CR-0.05
      IF(ML.GE.10)GO TO 1
      CR=CR+0.05
      GO TO 7
    6 IF(CR.GE.10.)GO TO 7
      IF(ML.GE.10)GO TO 7
      CR=CR+0.1
      GO TO 1
    7 CRP(K,L)=CR
      CALL TABTS1(Y,CRP,VCDX,GSBV,DX,S,EM,F)
    8 CONTINUE
      SECS=GETCLK(I)/100.
      WRITE(3,9)SECS
    9 FORMAT('-COMPUTE TIME WAS',F7.1,' SECONDS.')
      CALL EXIT
      STOP
      END
```

```
      SUBROUTINE IZETS1(DX,K,DXK,S,SK,EM,N4,CI,Y,VI,Q,F,VCDX,GSBV)
      DIMENSION DX(27),S(27),EM(27),F(27),VCDX(27),GSBV(27)
      DX(K)=DXK
      S(K)=SK
      EM(K)=0.006*2**N4
      CI=SQRT(32.2*Y)
      VI=1.486/EM(K)*SQRT(SK)*Y**(2./3.)
      Q=VI*Y
      F(K)=VI/CI
      VCDX(K)=(VI+CI)/DXK
      GSBV(K)=32.2*SK/VI
      RETURN
      END
```



```
     0SUBROUTINE HEDTS1(L,CR,VCDX,K,GSBV,M,Y,DXK,SK,EM,F,VCTX,GSTV,FEMDT
     1,GSDT,DTX,CI,VI,Q)
      DIMENSION VCDX(27),GSBV(27),EM(27),F(27)
      COMMON C(9),V(9)
      GO TO (21,22),L
   21 DT=CR/(VCDX(K)+GSBV(K))
      IF(M.GT.1)GO TO 25
      WRITE(3,90)Y,DXK,SK,EM(K),F(K)
   90OFORMAT('-DEPTH =',F7.3,' FT, MESH LENGTH =',F9.2,' FT, SLOPE =',F8
     1.5,', MANNING N =',F6.3,', FROUDE NUMBER =',F5.2/'0 STOKER FINITE
     2DIFFERENCE SCHEME')
      GO TO 25
   22 DT=CR/VCDX(K)
      IF(M.GT.1)GO TO 25
      WRITE(3,91)
   91 FORMAT('0 PROPOSED FINITE DIFFERENCE SCHEME')
   25 VCTX=VCDX(K)*DT
      GSTV=GSBV(K)*DT
      FEMDT=32.2**(7./3.)*DT*EM(K)*EM(K)/(1.486*1.486)
      GSDT=32.2*SK*DT
      DTX=0.25*DT/DXK
      WRITE(3,93)DT,VCTX,GSTV,CR
   93OFORMAT('0   TIME STEP =',F7.2,' SECONDS, (V+C)DT/DX =',F5.2,', GSD
     1T/V =',F5.2,' C.R. =',F5.2)
      DO 26 I=1,9
      C(I)=CI
   26 V(I)=VI
      C(5)=1.00001*C(5)
      V(5)=32.2*Q/(C(5)*C(5))
      RETURN
      END
```

```
      SUBROUTINE STRTS1(DTX,ML,FEMDT,GSDT)
      COMMON C(9),V(9),CN(9),VN(9)
      DO 59 I=2,8
      GV=(V(I+1)-V(I-1)+2.*(2.*C(I)-C(I+1)-C(I-1)))*DTX
      GC=(2.*(C(I+1)-C(I-1))+2.*V(I)-V(I+1)-V(I-1))*DTX
      CN(I)=C(I)-C(I)*GV-V(I)*GC
      IF(CN(I).LE.0.)ML=1000
      FEMDT=SIGN(FEMDT,V(I))
   59 VN(I)=V(I)-2.*C(I)*GC-2.*V(I)*GV+GSDT-FEMDT*V(I)*V(I)/(C(I)**(8./3
     1.))
      RETURN
      END
```



```
      SUBROUTINE BNTTS1(DTX,ML,GSDT,FEMDT)
      COMMON C(9),V(9),CN(9),VN(9)
      DO 69 I=2,8
      GV=(V(I+1)-V(I-1)+2.*(2.*C(I)-C(I+1)-C(I-1)))*DTX
      GC=(2.*(C(I+1)-C(I-1))+2.*V(I)-V(I+1)-V(I-1))*DTX
      CT=(C(I)-V(I)*GC)/(1.+GV)
      IF(CT.GT.0.)GO TO 65
      ML=1000
      GO TO 69
   65 EKC=V(I)-2.*CT*GC+GSDT
      EKA=SIGN(FEMDT,EKC)/(CT**(8./3.))
      EKB=1.+2.*GV
      AC=4.*EKA*EKC
      GPS=EKB*EKB
      IF(AC.LT.0.2*GPS)GO TO 66
      VN(I)=(-EKB+SQRT(GPS+AC))/(2.*EKA)
      GO TO 69
   66 ST1=0.25*AC/GPS
      ST2=ST1*ST1
      ST3=ST2*ST1*2.
      ST4=ST3*ST1*2.23
      VN(I)=EKB*(ST1-ST2+ST3-ST4)/EKA
   69 CN(I)=CT
      RETURN
      END
```

```
      SUBROUTINE TESTS1(J,ML,CI,VI,Y,EM,K,SK,*)
      DIMENSION EM(27)
      COMMON C(9),V(9),CN(9),VN(9)
      IF(ML.GE.1000)GO TO 85
      DIFF3=DIFF2
      DIFF2=DIFF
      DIFF=0.
      DO 81 I=2,8
      DIFF=DIFF+ABS((CN(I)-C(I))/CI)
   81 DIFF=DIFF+ABS((VN(I)-V(I))/VI)
      DIFF=DIFF/14.
      IF(DIFF.GT.2.E-7*(J-5))GO TO 83
      IF(DIFF.GT.DIFF2)GO TO 83
      IF(2.*DIFF.GT.DIFF2+DIFF3)GO TO 83
      WRITE(3,96)DIFF
   960FORMAT(' STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE ='
     1,E9.2)
      GO TO 88
   83 IF(J.GT.95)GO TO 86
      DO 84 I=2,8
      C(I)=CN(I)
   84 V(I)=VN(I)
      J=J+1
      RETURN 1
   85 WRITE(3,97)
   97 FORMAT(' UNSTABLE SOLUTION INDICATED. NEGATIVE DEPTH ENCOUNTERED')
      GO TO 88
   86 ML=10
      WRITE(3,98)DIFF
   98 FORMAT(' UNSTABLE SOLUTION PROBABLE AS SOLUTION NOT CONVERGING. AV
     1ERAGE PROPORTIONAL CHANGE =',E9.2)
   88 YR=CN(5)*CN(5)/(32.2*Y)
      SR=VN(5)*EM(K)/(1.486*(YR*Y)**(2./3.))
      SR=SR*SR/SK
      WRITE(3,99)YR,SR,J,CI,(CN(I),I=2,8),CI,VI,(VN(I),I=2,8),VI
   990FORMAT('    TYPICAL DEPTH/INITIAL DEPTH =',F7.4,', FRICTION SLOPE/
     1BED SLOPE =',F7.4,', NO. OF CYCLES =',I3/'    C =',9E14.7/'    V =
     2',9E14.7)
      RETURN
      END
```

```
      SUBROUTINE TABTS1(Y,CRP,VCDX,GSBV,DX,S,EM,F)
      DIMENSION CRP(27,2),VCDX(27),GSBV(27),DX(27),S(27),EM(27),F(27)
      WRITE(3,90)Y
   90 FORMAT('1SUMMARY OF STABILITY TEST RESULTS FOR UNIFORM FLOW AT DEP
     1TH OF',F7.3,' FEET'/'0    MESH    SLOPE  MANNING  FROUDE  DT2   GSD
     2X  SCHEME  TIME  (V+C)DT  GSDT  C.R.'/'+',35X,'___  ______',15X,'_
     3______  ____'/'    LENGTH',13X,'N',6X,'NO.    DT1  V(V+C)   NO.   S
     4TEP     DX      V'/'  DX(FEET)    S',6X,'EM',7X,'F',19X,'L   DT(SEC
     5S)')
      DO 99 K=1,27
      DT1=CRP(K,1)/(VCDX(K)+GSBV(K))
      DT2=CRP(K,2)/VCDX(K)
      DT2B1=DT2/DT1
      GSXBV=GSBV(K)/VCDX(K)
      VCTX1=VCDX(K)*DT1
      VCTX2=VCDX(K)*DT2
      GSTV1=GSBV(K)*DT1
      GSTV2=GSBV(K)*DT2
   99 WRITE(3,98)DX(K),S(K),EM(K),F(K),DT2B1,GSXBV,DT1,VCTX1,GSTV1,CRP(K
     1,1),DT2,VCTX2,GSTV2,CRP(K,2)
   98 FORMAT(' ',F8.2,F8.4,F8.3,F8.2,F6.2,F7.2,'     1',F10.2,2F7.2,F6.2/
     150X,'2',F10.2,2F7.2,F6.2)
      WRITE(3,97)
   97 FORMAT('0THESE RESULTS REFER TO THE CRITICAL TIME STEPS AT WHICH
     1INSTABILITY JUST EMERGED')
      RETURN
      END
```

# APPENDIX E. RESULTS OF STABILITY EXPERIMENTS

SUMMARY OF STABILITY TEST RESULTS FOR UNIFORM FLOW AT DEPTH OF 0.040 FEET

| MESH LENGTH DX(FEET) | SLOPE S | MANNING N EM | FROUDE NO. F | DT2/DT1 | GSDX/V(V+C) | SCHEME NO. L | TIME STEP DT(SECS) | (V+C)DT/DX | GSDT/V | C.R. |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.62 | 0.0001 | 0.012 | 0.13 | 1.01 | 0.01 | 1 | 0.67 | 1.39 | 0.01 | 1.40 |
| | | | | | | 2 | 0.68 | 1.40 | 0.02 | 1.40 |
| 0.62 | 0.0001 | 0.024 | 0.06 | 1.10 | 0.02 | 1 | 0.65 | 1.27 | 0.03 | 1.30 |
| | | | | | | 2 | 0.72 | 1.40 | 0.03 | 1.40 |
| 0.62 | 0.0001 | 0.048 | 0.03 | 0.97 | 0.05 | 1 | 0.66 | 1.24 | 0.06 | 1.30 |
| | | | | | | 2 | 0.63 | 1.20 | 0.06 | 1.20 |
| 0.62 | 0.0010 | 0.012 | 0.40 | 1.09 | 0.03 | 1 | 0.60 | 1.56 | 0.04 | 1.60 |
| | | | | | | 2 | 0.66 | 1.70 | 0.05 | 1.70 |
| 0.62 | 0.0010 | 0.024 | 0.20 | 1.14 | 0.06 | 1 | 0.60 | 1.32 | 0.08 | 1.40 |
| | | | | | | 2 | 0.68 | 1.50 | 0.10 | 1.50 |
| 0.62 | 0.0010 | 0.048 | 0.10 | 1.23 | 0.14 | 1 | 0.56 | 1.14 | 0.16 | 1.30 |
| | | | | | | 2 | 0.69 | 1.40 | 0.19 | 1.40 |
| 0.62 | 0.0100 | 0.012 | 1.28 | 1.05 | 0.05 | 1 | 0.25 | 1.04 | 0.06 | 1.10 |
| | | | | | | 2 | 0.26 | 1.10 | 0.06 | 1.10 |
| 0.62 | 0.0100 | 0.024 | 0.64 | 1.22 | 0.15 | 1 | 0.49 | 1.48 | 0.22 | 1.70 |
| | | | | | | 2 | 0.60 | 1.80 | 0.27 | 1.80 |
| 0.62 | 0.0100 | 0.048 | 0.32 | 1.56 | 0.37 | 1 | 0.42 | 1.02 | 0.38 | 1.40 |
| | | | | | | 2 | 0.66 | 1.60 | 0.59 | 1.60 |
| 1.96 | 0.0001 | 0.012 | 0.13 | 1.03 | 0.03 | 1 | 2.07 | 1.35 | 0.05 | 1.40 |
| | | | | | | 2 | 2.14 | 1.40 | 0.05 | 1.40 |
| 1.96 | 0.0001 | 0.024 | 0.06 | 1.15 | 0.07 | 1 | 1.96 | 1.21 | 0.09 | 1.30 |
| | | | | | | 2 | 2.27 | 1.40 | 0.10 | 1.40 |
| 1.96 | 0.0001 | 0.048 | 0.03 | 1.36 | 0.15 | 1 | 1.60 | 0.96 | 0.14 | 1.10 |
| | | | | | | 2 | 2.17 | 1.30 | 0.19 | 1.30 |
| 1.96 | 0.0010 | 0.012 | 0.40 | 1.09 | 0.09 | 1 | 1.81 | 1.47 | 0.13 | 1.60 |
| | | | | | | 2 | 1.96 | 1.60 | 0.14 | 1.60 |
| 1.96 | 0.0010 | 0.024 | 0.20 | 1.39 | 0.20 | 1 | 1.55 | 1.08 | 0.22 | 1.30 |
| | | | | | | 2 | 2.15 | 1.50 | 0.30 | 1.50 |
| 1.96 | 0.0010 | 0.048 | 0.10 | 1.68 | 0.44 | 1 | 1.30 | 0.83 | 0.37 | 1.20 |
| | | | | | | 2 | 2.19 | 1.40 | 0.62 | 1.40 |
| 1.96 | 0.0100 | 0.012 | 1.28 | 1.05 | 0.17 | 1 | 1.23 | 1.63 | 0.27 | 1.90 |
| | | | | | | 2 | 1.29 | 1.70 | 0.29 | 1.70 |
| 1.96 | 0.0100 | 0.024 | 0.64 | 1.74 | 0.47 | 1 | 1.15 | 1.09 | 0.51 | 1.60 |
| | | | | | | 2 | 2.00 | 1.90 | 0.89 | 1.90 |
| 1.96 | 0.0100 | 0.048 | 0.32 | 2.88 | 1.16 | 1 | 0.73 | 0.56 | 0.64 | 1.20 |
| | | | | | | 2 | 2.09 | 1.60 | 1.86 | 1.60 |
| 6.19 | 0.0001 | 0.012 | 0.13 | 1.28 | 0.11 | 1 | 5.67 | 1.17 | 0.13 | 1.30 |
| | | | | | | 2 | 7.25 | 1.50 | 0.16 | 1.50 |
| 6.19 | 0.0001 | 0.024 | 0.06 | 1.33 | 0.23 | 1 | 5.01 | 0.98 | 0.22 | 1.20 |
| | | | | | | 2 | 6.66 | 1.30 | 0.30 | 1.30 |
| 6.19 | 0.0001 | 0.048 | 0.03 | 1.74 | 0.47 | 1 | 3.95 | 0.75 | 0.35 | 1.10 |
| | | | | | | 2 | 6.87 | 1.30 | 0.61 | 1.30 |
| 6.19 | 0.0010 | 0.012 | 0.40 | 1.44 | 0.27 | 1 | 4.58 | 1.18 | 0.32 | 1.50 |
| | | | | | | 2 | 6.60 | 1.70 | 0.46 | 1.70 |
| 6.19 | 0.0010 | 0.024 | 0.20 | 2.05 | 0.64 | 1 | 3.32 | 0.73 | 0.47 | 1.20 |
| | | | | | | 2 | 6.80 | 1.50 | 0.96 | 1.50 |
| 6.19 | 0.0010 | 0.048 | 0.10 | 2.83 | 1.39 | 1 | 2.28 | 0.46 | 0.64 | 1.10 |
| | | | | | | 2 | 6.44 | 1.30 | 1.81 | 1.30 |
| 6.19 | 0.0100 | 0.012 | 1.28 | 1.46 | 0.53 | 1 | 3.28 | 1.37 | 0.73 | 2.10 |
| | | | | | | 2 | 4.79 | 2.00 | 1.06 | 2.00 |
| 6.19 | 0.0100 | 0.024 | 0.64 | 3.62 | 1.48 | 1 | 1.74 | 0.52 | 0.78 | 1.30 |
| | | | | | | 2 | 6.32 | 1.90 | 2.81 | 1.90 |
| 6.19 | 0.0100 | 0.048 | 0.32 | 6.37 | 3.67 | 1 | 0.97 | 0.24 | 0.86 | 1.10 |
| | | | | | | 2 | 6.20 | 1.50 | 5.51 | 1.50 |

THESE RESULTS REFER TO THE CRITICAL TIME STEPS AT WHICH INSTABILITY JUST EMERGED

SUMMARY OF STABILITY TEST RESULTS FOR UNIFORM FLOW AT DEPTH OF 1.000 FEET

| MESH LENGTH DX(FEET) | SLOPE S | MANNING N EM | FROUDE NO. F | DT2/DT1 | GSDX/V(V+C) | SCHEME NO. L | TIME STEP DT(SECS) | (V+C)DT/DX | GSDT/V | C.R. |
|---|---|---|---|---|---|---|---|---|---|---|
| 31.62 | 0.0001 | 0.012 | 0.22 | 1.01 | 0.01 | 1 | 7.23 | 1.58 | 0.02 | 1.60 |
| | | | | | | 2 | 7.32 | 1.60 | 0.02 | 1.60 |
| 31.62 | 0.0001 | 0.024 | 0.11 | 0.95 | 0.03 | 1 | 6.86 | 1.36 | 0.04 | 1.40 |
| | | | | | | 2 | 6.53 | 1.30 | 0.03 | 1.30 |
| 31.62 | 0.0001 | 0.048 | 0.05 | 1.05 | 0.05 | 1 | 6.51 | 1.23 | 0.07 | 1.30 |
| | | | | | | 2 | 6.87 | 1.30 | 0.07 | 1.30 |
| 31.62 | 0.0010 | 0.012 | 0.69 | 0.97 | 0.03 | 1 | 5.78 | 1.75 | 0.05 | 1.80 |
| | | | | | | 2 | 5.61 | 1.70 | 0.05 | 1.70 |
| 31.62 | 0.0010 | 0.024 | 0.35 | 1.13 | 0.07 | 1 | 6.21 | 1.50 | 0.10 | 1.60 |
| | | | | | | 2 | 7.04 | 1.70 | 0.12 | 1.70 |
| 31.62 | 0.0010 | 0.048 | 0.17 | 1.33 | 0.16 | 1 | 5.34 | 1.12 | 0.18 | 1.30 |
| | | | | | | 2 | 7.13 | 1.50 | 0.23 | 1.50 |
| 31.62 | 0.0100 | 0.012 | 2.18 | 7.67 | 0.05 | 1 | 0.25 | 0.14 | 0.01 | 0.15 |
| | | | | | | 2 | 1.93 | 1.10 | 0.05 | 1.10 |
| 31.62 | 0.0100 | 0.024 | 1.09 | 1.08 | 0.14 | 1 | 4.45 | 1.67 | 0.23 | 1.90 |
| | | | | | | 2 | 4.80 | 1.80 | 0.25 | 1.80 |
| 31.62 | 0.0100 | 0.048 | 0.55 | 1.74 | 0.38 | 1 | 3.93 | 1.09 | 0.41 | 1.50 |
| | | | | | | 2 | 6.85 | 1.90 | 0.71 | 1.90 |
| 100.00 | 0.0001 | 0.012 | 0.22 | 1.04 | 0.04 | 1 | 20.91 | 1.45 | 0.05 | 1.50 |
| | | | | | | 2 | 21.70 | 1.50 | 0.06 | 1.50 |
| 100.00 | 0.0001 | 0.024 | 0.11 | 1.17 | 0.08 | 1 | 19.08 | 1.20 | 0.10 | 1.30 |
| | | | | | | 2 | 22.24 | 1.40 | 0.12 | 1.40 |
| 100.00 | 0.0001 | 0.048 | 0.05 | 1.27 | 0.17 | 1 | 17.08 | 1.02 | 0.18 | 1.20 |
| | | | | | | 2 | 21.72 | 1.30 | 0.23 | 1.30 |
| 100.00 | 0.0010 | 0.012 | 0.69 | 1.09 | 0.09 | 1 | 16.33 | 1.57 | 0.13 | 1.70 |
| | | | | | | 2 | 17.73 | 1.70 | 0.15 | 1.70 |
| 100.00 | 0.0010 | 0.024 | 0.35 | 1.38 | 0.22 | 1 | 16.17 | 1.23 | 0.27 | 1.50 |
| | | | | | | 2 | 22.27 | 1.70 | 0.37 | 1.70 |
| 100.00 | 0.0010 | 0.048 | 0.17 | 1.74 | 0.49 | 1 | 12.07 | 0.80 | 0.40 | 1.20 |
| | | | | | | 2 | 21.04 | 1.40 | 0.69 | 1.40 |
| 100.00 | 0.0100 | 0.012 | 2.18 | 1.63 | 0.14 | 1 | 1.69 | 0.31 | 0.04 | 0.35 |
| | | | | | | 2 | 2.77 | 0.50 | 0.07 | 0.50 |
| 100.00 | 0.0100 | 0.024 | 1.09 | 1.51 | 0.44 | 1 | 11.13 | 1.32 | 0.58 | 1.90 |
| | | | | | | 2 | 16.85 | 2.00 | 0.88 | 2.00 |
| 100.00 | 0.0100 | 0.048 | 0.55 | 3.03 | 1.19 | 1 | 6.78 | 0.59 | 0.71 | 1.30 |
| | | | | | | 2 | 20.52 | 1.80 | 2.13 | 1.80 |
| 316.23 | 0.0001 | 0.012 | 0.22 | 1.20 | 0.12 | 1 | 57.23 | 1.25 | 0.15 | 1.40 |
| | | | | | | 2 | 68.62 | 1.50 | 0.18 | 1.50 |
| 316.23 | 0.0001 | 0.024 | 0.11 | 1.47 | 0.26 | 1 | 47.80 | 0.95 | 0.25 | 1.20 |
| | | | | | | 2 | 70.34 | 1.40 | 0.37 | 1.40 |
| 316.23 | 0.0001 | 0.048 | 0.05 | 1.68 | 0.55 | 1 | 40.92 | 0.77 | 0.43 | 1.20 |
| | | | | | | 2 | 68.70 | 1.30 | 0.71 | 1.30 |
| 316.23 | 0.0010 | 0.012 | 0.69 | 1.50 | 0.27 | 1 | 44.10 | 1.34 | 0.36 | 1.70 |
| | | | | | | 2 | 65.95 | 2.00 | 0.54 | 2.00 |
| 316.23 | 0.0010 | 0.024 | 0.35 | 2.07 | 0.68 | 1 | 32.03 | 0.77 | 0.53 | 1.30 |
| | | | | | | 2 | 66.29 | 1.60 | 1.09 | 1.60 |
| 316.23 | 0.0010 | 0.048 | 0.17 | 3.26 | 1.56 | 1 | 20.40 | 0.43 | 0.67 | 1.10 |
| | | | | | | 2 | 66.54 | 1.40 | 2.19 | 1.40 |
| 316.23 | 0.0100 | 0.012 | 2.18 | 1.46 | 0.46 | 1 | 18.05 | 1.03 | 0.47 | 1.50 |
| | | | | | | 2 | 26.27 | 1.50 | 0.68 | 1.50 |
| 316.23 | 0.0100 | 0.024 | 1.09 | 3.13 | 1.39 | 1 | 17.87 | 0.67 | 0.93 | 1.60 |
| | | | | | | 2 | 55.96 | 2.10 | 2.91 | 2.10 |
| 316.23 | 0.0100 | 0.048 | 0.55 | 7.13 | 3.75 | 1 | 9.11 | 0.25 | 0.95 | 1.20 |
| | | | | | | 2 | 64.90 | 1.80 | 6.75 | 1.80 |

THESE RESULTS REFER TO THE CRITICAL TIME STEPS AT WHICH INSTABILITY JUST EMERGED

SUMMARY OF STABILITY TEST RESULTS FOR UNIFORM FLOW AT DEPTH OF 25.000 FEET

| MESH LENGTH DX(FEET) | SLOPE S | MANNING N FM | FROUDE NO. F | DT2/DT1 | GSDX/V(V+C) | SCHEME NO. L | TIME STEP DT(SECS) | (V+C)DT/DX | GSDT/V | C.R. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1616.56 | 0.0001 | 0.012 | 0.37 | 1.01 | 0.01 | 1 | 65.56 | 1.58 | 0.02 | 1.60 |
| | | | | | | 2 | 66.39 | 1.60 | 0.02 | 1.60 |
| 1616.56 | 0.0001 | 0.024 | 0.19 | 1.10 | 0.03 | 1 | 65.32 | 1.36 | 0.04 | 1.40 |
| | | | | | | 2 | 72.03 | 1.50 | 0.04 | 1.50 |
| 1616.56 | 0.0001 | 0.048 | 0.09 | 1.15 | 0.06 | 1 | 63.71 | 1.22 | 0.08 | 1.30 |
| | | | | | | 2 | 72.96 | 1.40 | 0.09 | 1.40 |
| 1616.56 | 0.0010 | 0.012 | 1.18 | 1.03 | 0.03 | 1 | 28.04 | 1.07 | 0.03 | 1.10 |
| | | | | | | 2 | 28.75 | 1.10 | 0.03 | 1.10 |
| 1616.56 | 0.0010 | 0.024 | 0.59 | 1.13 | 0.07 | 1 | 56.99 | 1.59 | 0.11 | 1.70 |
| | | | | | | 2 | 64.50 | 1.80 | 0.12 | 1.80 |
| 1616.56 | 0.0010 | 0.048 | 0.30 | 1.25 | 0.17 | 1 | 56.44 | 1.28 | 0.22 | 1.50 |
| | | | | | | 2 | 70.39 | 1.60 | 0.27 | 1.60 |
| 1616.56 | 0.0100 | 0.012 | 3.73 | 3.11 | 0.04 | 1 | 0.58 | 0.05 | 0.00 | 0.05 |
| | | | | | | 2 | 1.81 | 0.15 | 0.01 | 0.15 |
| 1616.56 | 0.0100 | 0.024 | 1.87 | 1.54 | 0.12 | 1 | 7.09 | 0.36 | 0.04 | 0.40 |
| | | | | | | 2 | 10.93 | 0.55 | 0.07 | 0.55 |
| 1616.56 | 0.0100 | 0.048 | 0.93 | 1.59 | 0.36 | 1 | 39.05 | 1.32 | 0.48 | 1.80 |
| | | | | | | 2 | 61.90 | 2.10 | 0.75 | 2.10 |
| 5112.00 | 0.0001 | 0.012 | 0.37 | 1.04 | 0.04 | 1 | 201.88 | 1.54 | 0.06 | 1.60 |
| | | | | | | 2 | 209.94 | 1.60 | 0.06 | 1.60 |
| 5112.00 | 0.0001 | 0.024 | 0.19 | 1.17 | 0.09 | 1 | 194.61 | 1.28 | 0.12 | 1.40 |
| | | | | | | 2 | 227.76 | 1.50 | 0.14 | 1.50 |
| 5112.00 | 0.0001 | 0.048 | 0.09 | 1.40 | 0.20 | 1 | 164.73 | 1.00 | 0.20 | 1.20 |
| | | | | | | 2 | 230.72 | 1.40 | 0.28 | 1.40 |
| 5112.00 | 0.0010 | 0.012 | 1.18 | 1.08 | 0.08 | 1 | 130.15 | 1.57 | 0.13 | 1.70 |
| | | | | | | 2 | 140.50 | 1.70 | 0.14 | 1.70 |
| 5112.00 | 0.0010 | 0.024 | 0.59 | 1.36 | 0.22 | 1 | 158.16 | 1.40 | 0.30 | 1.70 |
| | | | | | | 2 | 215.30 | 1.90 | 0.41 | 1.90 |
| 5112.00 | 0.0010 | 0.048 | 0.30 | 1.89 | 0.54 | 1 | 117.81 | 0.85 | 0.45 | 1.30 |
| | | | | | | 2 | 222.61 | 1.60 | 0.86 | 1.60 |
| 5112.00 | 0.0100 | 0.012 | 3.73 | 1.12 | 0.12 | 1 | 6.83 | 0.18 | 0.02 | 0.20 |
| | | | | | | 2 | 7.62 | 0.20 | 0.02 | 0.20 |
| 5112.00 | 0.0100 | 0.024 | 1.87 | 1.18 | 0.38 | 1 | 90.96 | 1.45 | 0.55 | 2.00 |
| | | | | | | 2 | 106.88 | 1.70 | 0.65 | 1.70 |
| 5112.00 | 0.0100 | 0.048 | 0.93 | 2.93 | 1.13 | 1 | 69.89 | 0.75 | 0.85 | 1.60 |
| | | | | | | 2 | 205.07 | 2.20 | 2.49 | 2.20 |
| 16165.58 | 0.0001 | 0.012 | 0.37 | 1.20 | 0.13 | 1 | 589.49 | 1.42 | 0.18 | 1.60 |
| | | | | | | 2 | 705.37 | 1.70 | 0.21 | 1.70 |
| 16165.58 | 0.0001 | 0.024 | 0.19 | 1.49 | 0.29 | 1 | 483.12 | 1.01 | 0.29 | 1.30 |
| | | | | | | 2 | 720.26 | 1.50 | 0.44 | 1.50 |
| 16165.58 | 0.0001 | 0.048 | 0.09 | 1.91 | 0.63 | 1 | 382.73 | 0.73 | 0.47 | 1.20 |
| | | | | | | 2 | 729.60 | 1.40 | 0.89 | 1.40 |
| 16165.58 | 0.0010 | 0.012 | 1.18 | 1.25 | 0.25 | 1 | 396.83 | 1.52 | 0.38 | 1.90 |
| | | | | | | 2 | 496.57 | 1.90 | 0.48 | 1.90 |
| 16165.58 | 0.0010 | 0.024 | 0.59 | 2.14 | 0.69 | 1 | 318.19 | 0.89 | 0.61 | 1.50 |
| | | | | | | 2 | 680.84 | 1.90 | 1.31 | 1.90 |
| 16165.58 | 0.0010 | 0.048 | 0.30 | 3.37 | 1.69 | 1 | 196.08 | 0.45 | 0.75 | 1.20 |
| | | | | | | 2 | 659.95 | 1.50 | 2.54 | 1.50 |
| 16165.58 | 0.0100 | 0.012 | 3.73 | 2.34 | 0.37 | 1 | 61.70 | 0.51 | 0.19 | 0.70 |
| | | | | | | 2 | 144.50 | 1.20 | 0.44 | 1.20 |
| 16165.58 | 0.0100 | 0.024 | 1.87 | 1.86 | 1.21 | 1 | 170.98 | 0.86 | 1.04 | 1.90 |
| | | | | | | 2 | 318.10 | 1.60 | 1.93 | 1.60 |
| 16165.58 | 0.0100 | 0.048 | 0.93 | 7.76 | 3.59 | 1 | 83.56 | 0.28 | 1.02 | 1.30 |
| | | | | | | 2 | 648.49 | 2.20 | 7.89 | 2.20 |

THESE RESULTS REFER TO THE CRITICAL TIME STEPS AT WHICH INSTABILITY JUST EMERGED

```
DEPTH = 25.000 FT, MESH LENGTH = 16165.58 FT, SLOPE = 0.00100, MANNING N = 0.048, FROUDE NUMBER = 0.30

 STOKER FINITE DIFFERENCE SCHEME

   TIME STEP = 163.40 SECONDS, (V+C)DT/DX = 0.37, GSDT/V = 0.63 C.R. = 1.00
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.68E-06
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES =  9
   C = 0.2837251E 02 0.2837250E 02 0.2837248E 02 0.2837247E 02 0.2837247E 02 0.2837247E 02 0.2837248E 02 0.2837248E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370229E 01 0.8370230E 01 0.8370222E 01 0.8370214E 01 0.8370210E 01 0.8370230E 01 0.8370219E 01 0.8370221E 01

   TIME STEP = 179.74 SECONDS, (V+C)DT/DX = 0.41, GSDT/V = 0.69 C.R. = 1.10
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.16E-05
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES = 14
   C = 0.2837251E 02 0.2837248E 02 0.2837247E 02 0.2837245E 02 0.2837244E 02 0.2837242E 02 0.2837241E 02 0.2837239E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370236E 01 0.8370213E 01 0.8370234E 01 0.8370194E 01 0.8370215E 01 0.8370174E 01 0.8370187E 01 0.8370221E 01

   TIME STEP = 196.08 SECONDS, (V+C)DT/DX = 0.45, GSDT/V = 0.75 C.R. = 1.20
UNSTABLE SOLUTION PROBABLE AS SOLUTION NOT CONVERGING. AVERAGE PROPORTIONAL CHANGE = 0.41E 00
   TYPICAL DEPTH/INITIAL DEPTH = 0.9235, FRICTION SLOPE/BED SLOPE = 0.1317, NO. OF CYCLES = 96
   C = 0.2837251E 02 0.2828128E 02 0.2814386E 02 0.2800679E 02 0.2726572E 02 0.2735182E 02 0.2659567E 02 0.2748044E 02 0.2837251E 02
   V = 0.8370221E 01 0.9771767E 01 0.5128008E 01 0.1133042E 02 0.2880518E 01 0.1121786E 02 0.2598950E 01 0.1033370E 02 0.8370221E 01

 PROPOSED FINITE DIFFERENCE SCHEME

   TIME STEP = 439.97 SECONDS, (V+C)DT/DX = 1.00, GSDT/V = 1.69 C.R. = 1.00
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.32E-06
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES =  7
   C = 0.2837251E 02 0.2837250E 02 0.2837250E 02 0.2837250E 02 0.2837248E 02 0.2837251E 02 0.2837253E 02 0.2837253E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370219E 01 0.8370215E 01 0.8370215E 01 0.8370208E 01 0.8370221E 01 0.8370228E 01 0.8370234E 01 0.8370221E 01

   TIME STEP = 483.96 SECONDS, (V+C)DT/DX = 1.10, GSDT/V = 1.86 C.R. = 1.10
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.28E-06
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES =  7
   C = 0.2837251E 02 0.2837250E 02 0.2837250E 02 0.2837248E 02 0.2837251E 02 0.2837253E 02 0.2837254E 02 0.2837253E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370218E 01 0.8370214E 01 0.8370210E 01 0.8370219E 01 0.8370230E 01 0.8370246E 01 0.8370236E 01 0.8370221E 01

   TIME STEP = 527.96 SECONDS, (V+C)DT/DX = 1.20, GSDT/V = 2.03 C.R. = 1.20
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.23E-06
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES =  7
   C = 0.2837251E 02 0.2837251E 02 0.2837251E 02 0.2837251E 02 0.2837251E 02 0.2837251E 02 0.2837251E 02 0.2837253E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370230E 01 0.8370226E 01 0.8370224E 01 0.8370230E 01 0.8370226E 01 0.8370224E 01 0.8370228E 01 0.8370221E 01

   TIME STEP = 571.96 SECONDS, (V+C)DT/DX = 1.30, GSDT/V = 2.20 C.R. = 1.30
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.57E-06
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES =  8
   C = 0.2837251E 02 0.2837250E 02 0.2837248E 02 0.2837248E 02 0.2837250E 02 0.2837251E 02 0.2837250E 02 0.2837251E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370222E 01 0.8370222E 01 0.8370216E 01 0.8370216E 01 0.8370232E 01 0.8370222E 01 0.8370223E 01 0.8370221E 01

   TIME STEP = 615.95 SECONDS, (V+C)DT/DX = 1.40, GSDT/V = 2.37 C.R. = 1.40
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.38E-06
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES =  8
   C = 0.2837251E 02 0.2837251E 02 0.2837251E 02 0.2837250E 02 0.2837250E 02 0.2837250E 02 0.2837249E 02 0.2837251E 02 0.2837251E 02
   V = 0.8370221E 01 0.8370231E 01 0.8370234E 01 0.8370219E 01 0.8370214E 01 0.8370219E 01 0.8370213E 01 0.8370223E 01 0.8370221E 01

   TIME STEP = 659.95 SECONDS, (V+C)DT/DX = 1.50, GSDT/V = 2.54 C.R. = 1.50
UNSTABLE SOLUTION INDICATED. NEGATIVE DEPTH ENCOUNTERED
   TYPICAL DEPTH/INITIAL DEPTH = 0.7802, FRICTION SLOPE/BED SLOPE = 0.7950, NO. OF CYCLES = 73
   C = 0.2837251E 02 0.2900282E 02 0.2746252E 02 0.3737336E 02 0.2506149E 02 0.6630820E 02-0.1145956E 03-0.5737683E 01 0.2837251E 02
   V = 0.8370221E 01 0.8414732E 01 0.7992826E 01 0.9362473E 01 0.6325228E 01 0.4365603E 02 0.2073845E 01 0.2853652E 01 0.8370221E 01
```

```
DEPTH = 25.000 FT, MESH LENGTH = 16165.58 FT, SLOPE = 0.01000, MANNING N = 0.012, FROUDE NUMBER = 3.73

 STOKER FINITE DIFFERENCE SCHEME

   TIME STEP =   88.14 SECONDS, (V+C)DT/DX = 0.73, GSDT/V = 0.27 C.R. = 1.00
UNSTABLE SOLUTION INDICATED. NEGATIVE DEPTH ENCOUNTERED
   TYPICAL DEPTH/INITIAL DEPTH = 0.2760, FRICTION SLOPE/BED SLOPE = 0.7368, NO. OF CYCLES = 52
   C = 0.2837251E 02 0.2580713E 02 0.2321260E 02 0.3338556E 02 0.1490660E 02-0.1202122E 02 0.2638954E 02 0.2663782E 02 0.2837251E 02
   V = 0.1058758E 03 0.9896667E 02 0.1166870E 03 0.1211940E 03 0.3852878E 02-0.1171911E 03 0.1092079E 03 0.6569971E 02 0.1058758E 03

   TIME STEP =   83.73 SECONDS, (V+C)DT/DX = 0.70, GSDT/V = 0.25 C.R. = 0.95
UNSTABLE SOLUTION INDICATED. NEGATIVE DEPTH ENCOUNTERED
   TYPICAL DEPTH/INITIAL DEPTH = 0.6859, FRICTION SLOPE/BED SLOPE = 0.4016, NO. OF CYCLES = 60
   C = 0.2837251E 02 0.2670839E 02 0.2268410E 02 0.3203690E 02 0.2349818E 02-0.2325662E 01 0.1849565E 02 0.2033292E 02 0.2837251E 02
   V = 0.1058758E 03 0.9879810E 02 0.1130864E 03 0.1223879E 03 0.5218594E 02-0.1674128E 03 0.1082194E 03 0.7566756E 02 0.1058758E 03

   TIME STEP =   79.32 SECONDS, (V+C)DT/DX = 0.66, GSDT/V = 0.24 C.R. = 0.90
UNSTABLE SOLUTION INDICATED. NEGATIVE DEPTH ENCOUNTERED
   TYPICAL DEPTH/INITIAL DEPTH = 0.1031, FRICTION SLOPE/BED SLOPE =17.0869, NO. OF CYCLES = 72
   C = 0.2837251E 02 0.2829819E 02 0.3351855E 02 0.2257404E 02 0.9108444E 01-0.1445946E 04-0.1656403E 07-0.2273536E 04 0.2837251E 02
   V = 0.1058758E 03 0.1118876E 03 0.1013914E 03 0.5136165E 02 0.9620354E 02 0.3058760E 04 0.5398221E 09-0.4664277E 04 0.1058758E 03

   TIME STEP =   74.92 SECONDS, (V+C)DT/DX = 0.62, GSDT/V = 0.23 C.R. = 0.85
UNSTABLE SOLUTION INDICATED. NEGATIVE DEPTH ENCOUNTERED
   TYPICAL DEPTH/INITIAL DEPTH = 0.1800, FRICTION SLOPE/BED SLOPE = 8.8477, NO. OF CYCLES = 88
   C = 0.2837251E 02 0.2845900E 02 0.3279048E 02 0.2299748E 02 0.1203817E 02 0.1656903E 02-0.2467367E 03-0.2149042E 02 0.2837251E 02
   V = 0.1058758E 03 0.1109283E 03 0.1014121E 03 0.6272159E 02 0.1004069E 03 0.1381241E 02 0.7052133E 04-0.6967000E 02 0.1058758E 03

   TIME STEP =   70.51 SECONDS, (V+C)DT/DX = 0.59, GSDT/V = 0.21 C.R. = 0.80
UNSTABLE SOLUTION PROBABLE AS SOLUTION NOT CONVERGING. AVERAGE PROPORTIONAL CHANGE = 0.31E-01
   TYPICAL DEPTH/INITIAL DEPTH = 1.1350, FRICTION SLOPE/BED SLOPE = 0.8246, NO. OF CYCLES = 96
   C = 0.2837251E 02 0.2835471E 02 0.2740964E 02 0.2901920E 02 0.3022679E 02 0.2624600E 02 0.2606233E 02 0.3100503E 02 0.2837251E 02
   V = 0.1058758E 03 0.1048649E 03 0.1060885E 03 0.1090416E 03 0.1046123E 03 0.9810779E 02 0.1065053E 03 0.1105435E 03 0.1058758E 03

   TIME STEP =   66.10 SECONDS, (V+C)DT/DX = 0.55, GSDT/V = 0.20 C.R. = 0.75
UNSTABLE SOLUTION PROBABLE AS SOLUTION NOT CONVERGING. AVERAGE PROPORTIONAL CHANGE = 0.24E-02
   TYPICAL DEPTH/INITIAL DEPTH = 1.0037, FRICTION SLOPE/BED SLOPE = 1.0018, NO. OF CYCLES = 96
   C = 0.2837251E 02 0.2840485E 02 0.2837115E 02 0.2825632E 02 0.2842534E 02 0.2856302E 02 0.2823436E 02 0.2822214E 02 0.2837251E 02
   V = 0.1058758E 03 0.1058807E 03 0.1057240E 03 0.1058437E 03 0.1062330E 03 0.1058985E 03 0.1053592E 03 0.1059385E 03 0.1058758E 03

   TIME STEP =   61.70 SECONDS, (V+C)DT/DX = 0.51, GSDT/V = 0.19 C.R. = 0.70
UNSTABLE SOLUTION PROBABLE AS SOLUTION NOT CONVERGING. AVERAGE PROPORTIONAL CHANGE = 0.17E-03
   TYPICAL DEPTH/INITIAL DEPTH = 0.9990, FRICTION SLOPE/BED SLOPE = 1.0012, NO. OF CYCLES = 96
   C = 0.2837251E 02 0.2837160E 02 0.2837820E 02 0.2837144E 02 0.2835892E 02 0.2837395E 02 0.2838914E 02 0.2836233E 02 0.2837251E 02
   V = 0.1058758E 03 0.1058816E 03 0.1058783E 03 0.1058551E 03 0.1058703E 03 0.1059128E 03 0.1058801E 03 0.1058374E 03 0.1058758E 03

   TIME STEP =   57.29 SECONDS, (V+C)DT/DX = 0.48, GSDT/V = 0.17 C.R. = 0.65
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.56E-05
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES = 34
   C = 0.2837251E 02 0.2837250E 02 0.2837221E 02 0.2837242E 02 0.2837289E 02 0.2837227E 02 0.2837164E 02 0.2837268E 02 0.2837251E 02
   V = 0.1058758E 03 0.1058755E 03 0.1058755E 03 0.1058763E 03 0.1058759E 03 0.1058741E 03 0.1058746E 03 0.1058763E 03 0.1058758E 03

 PROPOSED FINITE DIFFERENCE SCHEME

   TIME STEP = 120.42 SECONDS, (V+C)DT/DX = 1.00, GSDT/V = 0.37 C.R. = 1.00
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.13E-05
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES = 12
   C = 0.2837251E 02 0.2837247E 02 0.2837244E 02 0.2837248E 02 0.2837251E 02 0.2837241E 02 0.2837245E 02 0.2837245E 02 0.2837251E 02
   V = 0.1058758E 03 0.1058759E 03 0.1058758E 03 0.1058758E 03 0.1058756E 03 0.1058756E 03 0.1058758E 03 0.1058759E 03 0.1058758E 03

   TIME STEP = 132.46 SECONDS, (V+C)DT/DX = 1.10, GSDT/V = 0.40 C.R. = 1.10
STABLE SOLUTION INDICATED. AVERAGE PROPORTIONAL CHANGE = 0.34E-05
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES = 22
   C = 0.2837251E 02 0.2837248E 02 0.2837241E 02 0.2837260E 02 0.2837239E 02 0.2837247E 02 0.2837256E 02 0.2837234E 02 0.2837251E 02
   V = 0.1058758E 03 0.1058758E 03 0.1058757E 03 0.1058761E 03 0.1058755E 03 0.1058757E 03 0.1058759E 03 0.1058754E 03 0.1058758E 03

   TIME STEP = 144.50 SECONDS, (V+C)DT/DX = 1.20, GSDT/V = 0.44 C.R. = 1.20
UNSTABLE SOLUTION PROBABLE AS SOLUTION NOT CONVERGING. AVERAGE PROPORTIONAL CHANGE = 0.52E-04
   TYPICAL DEPTH/INITIAL DEPTH = 1.0000, FRICTION SLOPE/BED SLOPE = 1.0000, NO. OF CYCLES = 96
   C = 0.2837251E 02 0.2837253E 02 0.2837248E 02 0.2837245E 02 0.2837265E 02 0.2837170E 02 0.2837314E 02 [illegible] 02 0.2837251E 02
   V = 0.1058758E 03 0.1058760E 03 0.1058759E 03 0.1058757E 03 0.1058762E 03 0.1058739E 03 0.1058784E 03 [illegible] 03 0.1058758E 03
```

## APPENDIX F. HAYAMI'S DIFFUSION SOLUTION

A recapitulation of the pertinent parts of the diffusion solution presented by Hayami (1951) is given here to complement the discussion in Chapter 8. By analysing flow in a rectangular channel of constant width, using the Chézy formula (2-39) for friction slope, Hayami obtained an equation which can be written

$$y_t + cy_s = Ky_{ss} \tag{F-1}$$

which is the same as (8-7), where c was defined as $\frac{3}{2}V$, corresponding to (8-4) with $m = \frac{1}{2}$, $k = 1$. However in defining the coefficient of $y_{ss}$ Hayami assumed that the fluctuations in width in a natural river may be treated as analagous to eddy diffusion. Thus, while assuming the mean channel was of constant width, Hayami allowed for storage in the wider parts of an irregular channel by introducing a "diffusion term" $Jy_{ss}$ to the R.H.S. of the equivalent of (8-7), so that in effect K in (F-1) was defined by

$$K = \frac{Vy}{2S_f} + J \tag{F-2}$$

(cf. (8-5) with $j = \frac{1}{2}$, $\cos\phi = 1$, $k = 1$).

Now initially $y(s,0) = Y_0$

and at the upper end $y(0,t > 0) = Y_0 + y_0 + F(t)$ where $\overline{F(t)} = 0$.

Hayami assumed the solution of (F-1) could be expressed as a functional series

$$y = (Y_0 + y_0)\left(1 + \frac{f_1}{Y_0 + y_0} + \frac{f_2}{(Y_0 + y_0)^2} + \ldots.\right)$$

where $f_1$, $f_2$ .... satisfy the conditions at $s = 0$:

$$f_1 = F(t); \quad f_2, f_3 \ldots. = 0$$

A first approximate solution is $y = Y_0 + y_0 + f_1(s,t)$ (F-3)

Substituting into (F-1)

$$(Y_0 + y_0)_t + f_{1t} + c(Y_0 + y_0)_s + cf_{1s} = K(Y_0 + y_0)_{ss} + Kf_{1ss}$$

Hayami assumed that the initial depth $Y_0$ and the average rise $y_0$ are constants, giving

$$f_{1t} + cf_{1s} = Kf_{1ss} \tag{F-4}$$

Boundary conditions for $t > 0$ are:

At $s = 0$ $f_1(0, t > 0) = F(t)$

At $s = s_1$ where $s_1$ is very large or else the point at which the channel empties into a reservoir

$$f_{1s} = 0$$

The initial condition is $f_1(s,0) = -y_0$

Hayami produced the solution

$$f_1 = -y_0 + \frac{s}{2\sqrt{\pi K}} \exp\left(\frac{cs}{2K}\right) \int_0^t \left[F(\lambda) + y_0\right] \frac{\exp\left[-\frac{c^2}{4K}(t-\lambda) - \frac{s^2}{4K(t-\lambda)}\right]}{(t-\lambda)^{3/2}} d\lambda \tag{F-5}$$

This simplified, for the case of a square step pulse such that $F(t) = 0$ for all $t$, to

$$f_1 = y_0 \left\{ \int_0^t \frac{s}{2\sqrt{\pi K}(t-\lambda)^{3/2}} \exp\left[-\left(\frac{c}{2}\sqrt{\frac{t-\lambda}{K}} - \frac{s}{2\sqrt{K(t-\lambda)}}\right)^2\right] d\lambda - 1 \right\} \tag{F-6}$$

Hayami did not present the derivation of (F-5), but we can show that (F-6) satisfies (F-4) and the given initial and boundary conditions by the following argument. For convenience, we define

$$\alpha = \frac{s}{2\sqrt{\pi K}\ (t-\lambda)^{3/2}}$$

$$\gamma = -\left[\frac{c}{2}\sqrt{\frac{t-\lambda}{K}} - \frac{s}{2\sqrt{K(t-\lambda)}}\right]^2$$

Differentiating under the integral sign, noting that a small increment in the limit t does not affect the integral, as when $\lambda$ tends to t the integrand tends to zero, we obtain

$$\left(\frac{f_1}{y_0}\right)_t = \int_0^t \alpha e^{\gamma}\left[-\frac{3}{2(t-\lambda)} - \frac{c^2}{4K} + \frac{s^2}{4K(t-\lambda)^2}\right]d\lambda \qquad \text{(F-7)}$$

$$\left(\frac{f_1}{y_0}\right)_s = \int_0^t \alpha e^{\gamma}\left[\frac{1}{s} + \frac{c}{2K} - \frac{s}{2K(t-\lambda)}\right]d\lambda \qquad \text{(F-8)}$$

$$\left(\frac{f_1}{y_0}\right)_{ss} = \int_0^t \alpha e^{\gamma}\left[\frac{c}{Ks} - \frac{3}{2K(t-\lambda)} + \frac{c^2}{4K^2} - \frac{cs}{2K^2(t-\lambda)} + \frac{s^2}{4K^2(t-\lambda)^2}\right]d\lambda \qquad \text{(F-9)}$$

It is therefore clear that for all s,t

$$\left(\frac{f_1}{y_0}\right)_t + c\left(\frac{f_1}{y_0}\right)_s - K\left(\frac{f_1}{y_0}\right)_{ss} = 0 \qquad \text{(F-10)}$$

Hence $f_1$ satisfies (F-4) if $y_0$ is constant.

It is now convenient to substitute $Z = \dfrac{s}{2\sqrt{K(t-\lambda)}}$ (F-11)

Note $\dfrac{dZ}{d\lambda} = \dfrac{s}{4K^{\frac{1}{2}}(t-\lambda)^{3/2}} = \dfrac{\alpha\sqrt{\pi}}{2}$

When $\lambda = 0$, $Z = \dfrac{s}{2\sqrt{Kt}}$ ; when $\lambda = t$, $Z = \infty$

(F-6) therefore becomes

$$f_1 = y_0\left\{\int_{\frac{s}{2\sqrt{Kt}}}^{\infty} \frac{2}{\sqrt{\pi}} \exp\left[-\left(\frac{cs}{4KZ} - Z\right)^2\right]dZ - 1\right\} \qquad \text{(F-12)}$$

Or if we introduce $G = \frac{cs}{2K}$ (F-13)

$$H = \frac{s}{2\sqrt{Kt}} \tag{F-14}$$

$$f_1 = y_0 \left\{ \frac{2}{\sqrt{\pi}} \int_H^{\infty} \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] dZ - 1 \right\} \tag{F-15}$$

Now by Abramowitz and Stegun (1965) Eq. 7.4.3,

$$\int_0^{\infty} \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] dZ = \frac{\sqrt{\pi}}{2} \tag{F-16}$$

Hence as $H \to 0$, i.e. as $t \to \infty$, $f_1 \to 0$ as we would expect.

This also shows that we may express (F-15) and hence (F-6) as

$$f_1 = \frac{-2y_0}{\sqrt{\pi}} \int_0^{H} \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] dZ \tag{F-17}$$

It is clear that $f_1(s,0) = -y_0$, consistent with the initial condition. Also as $H = 0$ for $s = 0$, $t > 0$ the upstream boundary condition is satisfied. Finally, we recast (F-8) using (F-11), (F-13) and (F-14) giving

$$\left(\frac{f_1}{y_0}\right)_s = \frac{2}{\sqrt{\pi}} \int_H^{\infty} (1 + G - 2Z^2) \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] \frac{dZ}{s} \tag{F-18}$$

Thus the downstream boundary condition is also satisfied for s large and $y_0$ constant. We have therefore verified Hayami's derivation.

Combining (F-3) with (F-17)

$$y = Y_0 + y_0 \left\{ 1 - \frac{2}{\sqrt{\pi}} \int_0^{H} \exp\left[-\left(\frac{G}{2Z} - Z\right)^2\right] dZ \right\} \tag{F-19}$$

This equation is therefore the solution of (F-1) satisfying the given initial and boundary conditions.